Multimodal Training of Maintenance and Assembly Skills Based on Augmented Reality

vom Fachbereich Informatik
der Technischen Universität Darmstadt
genehmigte

DISSERTATION

zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs (Dr.-Ing.)
von

Dipl.-Inform. Sabine Webel

geboren in Worms, Deutschland

Referenten der Arbeit:	Prof. Dr. techn. Dieter W. Fellner Technische Universität Darmstadt Prof. Dr.-Ing. Gerhard Hirzinger DLR (Deutsches Zentrum für Luft- und Raumfahrt e.V.)
Tag der Einreichung:	18/10/2011
Tag der mündlichen Prüfung:	07/12/2011

Darmstädter Dissertation
D 17
Darmstadt, 2012

Erklärung zur Dissertation

Hiermit versichere ich die vorliegende Dissertation selbständig nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die aus Quellen entnommen wurden, sind als solche kenntlich gemacht. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Darmstadt, den 18.10.2011 Sabine Webel

To my family

Acknowledgements

During the last years of working on this thesis I was greatly supported by the team of the Department of Virtual and Augmented Reality at the Fraunhofer IGD in Darmstadt. I am very grateful for a pleasant and motivating environment in which to work and the help and friendship of my colleagues. First of all, many thanks go to my PhD supervisor Prof. Dr. Dieter Fellner for his support. I'd also like to thank PD Dr. Arjan Kuijper for his assistance and advice when writing this thesis and Dr. Uli Bockholt for fruitful discussions. I wish to extend my sincere thanks to Prof. Dr. Gerhard Hirzinger for his assistance and advice.

In addition I'd like to thank the Department of Robotic Systems of the DLR Institute of Robotics and Mechatronics for providing the haptic device that I used in my work and useful studies, and I thank all partners of the SKILLS consortium for the valuable collaboration during the last years.

Finally, I express my sincere gratitude to my family for their persistence and motivating encouragement.

Abstract

The training of technicians in the acquisition of new maintenance and assembly tasks is an important factor in industry. As the complexity of these tasks can be enormous, the training of technicians to acquire the necessary skills to perform them efficiently is a challenging point. However, traditional training programs are usually highly theoretical and it is difficult for the trainees to transfer the acquired theoretical knowledge about the task to the real task conditions, or rather, to the physical performance of the task. In addition, traditional training programs are often expensive in terms of effort and cost.

Previous research has shown that Augmented Reality is a powerful technology to support training in the particular context of industrial service procedures, since instructions on how to perform the service tasks can be directly linked to the machine parts to be processed. Various approaches exist, in which the trainee is guided step-by-step through the maintenance task, but these systems act more as guiding systems than as training systems and focus only on the trainees' sensorimotor capabilities. Due to the increasing complexity of maintenance tasks, it is not sufficient to train the technicians' execution of these tasks, but rather to train the underlying skills—sensorimotor and cognitive—that are necessary for an efficient acquisition and performance of new maintenance operations.

All these facts lead to the need for efficient training systems for the training of maintenance and assembly skills which accelerate the technicians' learning and acquisition of new maintenance procedures. Furthermore, these systems should improve the adjustment of the training process to new training scenarios and enable the reuse of existing training material that has proven its worth. In this thesis a novel concept and platform for multimodal Augmented Reality-based training of maintenance and assembly skills is presented. This concept includes the identification of necessary sub-skills, the training of the involved skills, and the design of a training program for the training of maintenance and assembly skills. Since *procedural skills* are considered as the most important skills for maintenance and assembly operations, they are discussed in detail, as well as appropriate methods for improving them. We further show that the application of Augmented Reality technologies and the provision of multimodal feedback—and vibrotactile feedback in particular—have a great potential to enhance skill training in general. As a result, training strategies and specific *accelerators* for the training of maintenance and assembly skills in general and procedural skills in particular are elaborated. Here, accelerators are concrete methods used to implement the pursued training strategies.

Furthermore, a novel concept for displaying location-dependent information in Augmented Reality environments is introduced, which can compensate tracking imprecisions. In this concept, the pointer-content metaphor of annotating documents is transferred to Augmented Reality environments. As a result, *Adaptive Visual Aids* are defined which consist of a tracking-dependent pointer object and a tracking-independent content object, both providing an adaptable level and type of information. Thus,

the guidance level of Augmented Reality overlays in AR-based training applications can be easily controlled. Adaptive Visual Aids can be used to substitute traditional Augmented Reality overlays (i.e. overlays in form of 3D animations), which highly suffer from tracking inaccuracies.

The design of the multimodal AR-based training platform proposed in this thesis is not specific for the training of maintenance and assembly skills, but is a general design approach for multimodal training platforms. We further present an implementation of this platform based on the *X3D* ISO standard which provides features that are useful for the development of Augmented Reality environments. This standard-based implementation increases the sustainability and portability of the platform.

The implemented multimodal Augmented Reality-based platform for training of maintenance and assembly skills has been evaluated in industry and compared to traditional training methods. The results show that the developed training platform and the pursued training strategies are very well suited for the training of maintenance and assembly skills and enhance traditional training.

With the presented framework we have overcome the problems sketched above. We are cheap in terms of effort and costs for the training of maintenance and assembly skills and we improve its efficiency compared with traditional training.

Zusammenfassung

Das Training von Servicetechnikern in der Aneignung neuer Arbeitsvorgänge im Bereich Wartung und Montage ist in der Industrie von enormer Bedeutung. Aufgrund der steigenden Komplexität dieser Arbeitsvorgänge stellt das effiziente Training von Technikern im Erwerb der erforderlichen Fähigkeiten eine große Herausforderung dar. Es reicht nicht länger aus, Techniker nur in der physikalischen Ausführung dieser Aufgaben zu trainieren. Vielmehr ist es wichtig, auch die zugrunde liegenden Fähigkeiten zu adressieren. Das bedeutet, die Fähigkeiten, die für eine effiziente Aneignung und Ausführung der Aufgaben erforderlich sind, müssen in das Training einbezogen werden. Dies führt zu drei Hauptanforderungen, die Systeme zum Training von Fähigkeiten erfüllen müssen: Geschwindigkeit, Effizienz und Übertragbarkeit des Trainings. Um dies zu erreichen, müssen Trainingsprozeduren entwickelt werden, die den Lernprozess des Benutzers unterstützen und den Transfer der trainierten Fähigkeiten auf reale Arbeitssituationen erleichtern.

Traditionelles Training ist im Allgemeinen sehr theoretisch. Dadurch ist es für den Techniker schwierig, das erlernte theoretische Wissen über den Arbeitsvorgang auf die reale Ausführung des Arbeitsvorgangs zu übertragen. Ein weiterer kritischer Faktor ist der hohe Zeit- und Kostenaufwand, den traditionelles Training mit sich führt.

Bisherige Forschungsansätze haben bereits das große Potential von Augmented Reality (AR) als trainingsunterstützende Technologie, insbesondere im Bereich industrieller Wartungs- und Montageprozeduren, aufgezeigt. Mittels Augmented Reality können digitale Instruktionen direkt in der realen Umgebung verortet und somit direkt an reale Maschinenteile angeheftet werden. In den bisherigen Ansätzen stehen allerdings meist rein technologische Aspekte im Vordergrund. Des Weiteren fungieren existierende Systeme eher als digitale Benutzerhandbücher und unterstützen den Techniker vor allem bei der Ausführung der entsprechenden Aufgaben, anstatt bei der Aneignung der zugrunde liegenden—sensomotorischen und kognitiven—Fähigkeiten. Die Systeme führen den Benutzer Schritt für Schritt durch den Arbeitsvorgang, indem detaillierte 3D-Animationen der auszuführenden Aktion in Überlagerung mit dem realen Objekt visualisiert werden. Eine solch starke Führung des Benutzers mittels Augmented Reality birgt die potentielle Gefahr, dass der Benutzer für die Ausführung eines Arbeitsvorgangs eine Abhängigkeit von den bereitgestellten Augmented Reality-Elementen entwickelt, und die Aufgabe bei Nichtvorhandensein dieser Elemente nicht bewerkstelligen kann. Daher muss sich ein Augmented Reality-basiertes Trainingssystem deutlich von einem Augmented Reality-basierten Unterstützungs- bzw. Leitsystem unterscheiden: Es muss den Benutzer tatsächlich trainieren und die aktive Erkundung der Aufgabe forcieren, anstatt ihn lediglich zu dem gewünschten Ziel zu führen. Dies kann nur erreicht werden, indem sowohl sensomotorische als auch kognitive Aspekte in das Training einbezogen werden.

Eines der Hauptmerkmale für die Gestaltung von effizienten Trainingsprogrammen ist die gute Führung des Benutzers durch den zu trainierenden Arbeitsprozess. Wie bereits erwähnt, sind traditionelle Trainingsprogramme zum einen mit einem hohen Zeit- Kostenaufwand verbunden, zum anderen sind sie im Allgemeinen sehr theoretisch (nicht direkt am Objekt, sondern in Klassenräumen), was den Transfer der erlernten Trainingsaufgabe auf die reale Wartungs- oder Montageaufgabe erschwert und somit das Training zu einem langwierigen und demzufolge kostspieligen Prozess macht. Auch die Erstellung des Schulungsmaterials, zumeist Folienpräsentationen, die jeden Schritt umschreiben, ist sehr aufwendig und aufgabenspezifisch.

All dies zeigt die Notwendigkeit effizienter Trainingssysteme, die den Lern- bzw. Aneignungsprozess neuer Wartungs- und Montageaufgaben verbessern und beschleunigen. Diese Systeme sollen auch eine vereinfachte Anpassbarkeit des Systems an neue Trainingsszenarien und die Wiederverwendung existierender, bewährter Trainingsmaterialien ermöglichen. In dieser Arbeit wird ein neues Konzept für die Gestaltung von effizientem Training von Wartungs- und Montageprozessen unter Verwendung von Augmented Reality-Technologien vorgestellt. Dies beinhaltet unter anderem Visualisierungsparadigmen zum Präsentieren von Instruktionen und Informationen in multimodalen Augmented Reality-Trainingsumgebungen und ein Designkonzept für die Entwicklung einer multimodalen Trainingsplattform zur Umsetzung des Trainings.

Das vorgestellte Konzept berücksichtigt die Identifikation der erforderlichen sensomotorischen und kognitiven Fähigkeiten (so genannte *Sub-Skills* bzw. Unterkategorien von Fähigkeiten), das Training der involvierten Fähigkeiten und die Entwicklung eines Programms zum Training von Wartungs- und Montagefähigkeiten. In diesem Zusammenhang wird besonders auf die prozeduralen Fähigkeiten (so genannte *Procedural Skills*), die als wichtigste Sub-Skill betrachtet werden, eingegangen. Des Weiteren wird gezeigt, dass der Einsatz von Augmented Reality und das Bereitstellen von multimodalem Feedback—vor allem vibrotaktilem Feedback—großes Potential zum Training von Fähigkeiten mit sich bringen. Darauf basierend werden Trainingsstrategien und so genannte *Accelerators* für das Training von Wartungs- und Montagefähigkeiten im Allgemeinen und Procedural Skills im Speziellen erarbeitet. Dabei verweist der Ausdruck *Accelerators* auf Variablen bzw. konkrete Methoden, die in das Trainingsprogramm eingeführt werden, um die verfolgten Trainingsstrategien umzusetzen.

Darüber hinaus wird ein neues Konzept zur Visualisierung von ortsabhängigen Informationen in Augmented Reality-Umgebungen vorgestellt, durch welches Trackingungenauigkeiten kompensiert werden können. In diesem Konzept wird die zum Annotieren von Dokumenten typische *Zeiger-Inhalt* Metapher in Augmented Reality-Umgebungen übertragen. Als Ergebnis werden *Adaptive Visual Aids (AVAs)* definiert, die aus einem trackingabhängigen Zeigerobjekt und einem trackingunabhängigen Inhaltsobjekt bestehen. Informationsgehalt und Art der Information (Text, Bild, Video, 3D-Inhalte) von Zeiger- und Inhaltsobjekt sind regulierbar. Auf diese Weise kann die Stärke der Führung des Benutzers, die durch überlagerte 3D-Objekte in Augmented Reality bereitgestellt wird, einfach gesteuert werden. AVAs können auch benutzt werden, um traditionelle Augmented Reality-Einblendungen (d.h. 3D-Modelle und -Animationen), die sehr unter Trackingungenauigkeiten leiden, zu ersetzen. Außerdem können vorhandene Trainingsmaterialen, wie z.B. Bilder, Videos oder Beschreibungen in Form von Text, im Inhaltsobjekt einer AVA eingebunden werden.

Das Designkonzept der multimodalen Trainingsplattform, das in dieser Arbeit vorgestellt wird, ist nicht spezifisch für das Training von Wartungs- und Montagefähigkeiten entwickelt, sondern stellt vielmehr ein generelles Designkonzept für multimodale Trainingsplattformen dar. Es wird gezeigt, wie das entwickelte Plattformkonzept auf Basis des X3D ISO Standards implementiert werden kann. Dieser Standard wurde ausgewählt, da er Merkmale bereitstellt, die für die Entwicklung von Augmented Reality-Umgebungen nützlich sind. Eine solch standardbasierte Implementierung ist ein wichtiger Faktor zur Steigerung der Nachhaltigkeit und Portabilität der Plattform.

Die implementierte Plattform und die erarbeiteten Accelerators wurden in der Industrie evaluiert und den dort angewendeten traditionellen Trainingsmethoden gegenübergestellt. Das Ergebnis zeigt, dass das entwickelte Trainingssystem sehr gut zum Training von Wartungs- und Montagefähigkeiten geeignet ist, und dass dadurch traditionelles Training aufwertet werden kann. Mit dem vorgestellten Framework können die oben skizzierten Probleme bewältigt werden. Aufwand und Kosten für das Training von Wartungs- und Montagefähigkeiten werden reduziert, die Trainingseffizienz im Vergleich zu traditionellem Training wird gesteigert.

Kapitelübersicht

Introduction

Da die Komplexität von Wartungs- und Montagearbeiten im Bereich Industrie enorm sein kann und stetig wächst, steigt natürlich auch die Nachfrage nach geeigneten, effizienten Trainingssystemen zum Training dieser Aufgaben. Aus diesem Grund reicht eine reine Unterstützung der Techniker während der Ausführung der Wartungsarbeiten nicht länger aus, um die Effizienz des Trainings zu steigern. Vielmehr müssen die Fähigkeiten der Techniker, die zum schnellen Verständnis der Wartungsprozeduren und zur effizienten Aneignung neuer Wartungsvorgänge notwendig sind, trainiert werden. Um dies zu erreichen, müssen Trainingsprozeduren entwickelt werden, die den Lernprozess des Benutzers erleichtern und beschleunigen. Genauer gesagt, es müssen Elemente in das Training integriert werden, die den Benutzer beim Durchlaufen des folgenden Prozesses unterstützen: Beobachtung eines Modells, Erstellung geeigneter Repräsentationen von motorischen und verhaltensspezifischen Reaktionen, Überführung von speziellen Instruktionen in angemessene Aktionen, Erweiterung von einfachen Motorik- und Verhaltensprogrammen entsprechend neuer und komplexerer Anforderungen, und Transfer der Fähigkeiten in verschiedene Situationen.

Traditionelles Training ist im Allgemeinen sehr theoretisch. Dadurch ist es für den Techniker schwierig, das erlernte theoretische Wissen über den Arbeitsvorgang auf die reale Ausführung des Arbeitsvorgangs zu übertragen. Ein weiterer kritischer Faktor ist der hohe Zeit- und Kostenaufwand, den traditionelles Training mit sich führt. Daher besteht eine hohe Notwendigkeit zur Entwicklung von Trainingssystemen, die den gesamten Trainingsprozess, bestehend aus der Vorbereitung des Trainings (u.a. Erstellung der Trainingsmaterialien) und dem Training selbst, beschleunigen bzw. verbessern.

In dieser Arbeit wird ein Konzept vorgestellt, wie das Training von Wartungs- und Montagefähigkeiten durch die Verwendung von Augmented Reality-Technologien in multimodalen Trainingsumgebungen verbessert werden kann. Darüber hinaus werden eine standardbasierte Umsetzung des Konzeptes und die Evaluierung der resultierenden Trainingsplattform beschrieben.

Augmented Reality-Based and Multimodal Training Systems

Forschung im Bereich Design und Entwicklung von Trainingssystemen hat sich im Laufe der letzten Jahre enorm verstärkt. Dabei werden verschiedene Forschungsfelder abgedeckt, wie z.B. Training in den Bereichen Industrie, Chirurgie, Rehabilitation und Sport. So existieren auch Ansätze, die sich mit den Herausforderungen der Integration multimodaler Schnittstellen und fortgeschrittenen Trainingsumgebungen, wie zum Beispiel Augmented Reality-Umgebungen.

Betrachtet man die vorhandenen Ansätze so wird deutlich, dass existierende Systeme zum Training von Fähigkeiten sich entweder auf rein technologische Aspekte fokussieren, wie das Erfassen von Kräften und Bewegungen oder die Entwicklung von Simulationsmodellen und Rendering-Technologien, oder sie konzentrieren sich auf das Training der zugrunde liegenden sensomotorischen Fähigkeiten. Die meisten Ansätze führen den Benutzer sehr stark durch die Trainingsaufgabe und hemmen dadurch das aktive Eingebundensein des Benutzers in der Aneignung bzw. Verbesserung der Fähigkeit(en). Auf diese Weise erlangt der Benutzer die motorischen Befähigung zum Ausführen neuer Aufgaben, allerdings erwirbt er nicht die zugrunde liegenden Fähigkeiten, die zum Transfer der erlernten Trainingsaufgabe auf ähnliche Situationen und zur effizienten Aneignung neuer ähnlicher Aufgaben notwendig sind.

Um die Fähigkeiten zu trainieren, die komplexen Aufgaben (z.B. Wartungs- und Montageaufgaben, chirurgische Vorgänge) zugrunde liegen, müssen die kognitiven Aspekte dieser Aufgaben bzw. Fähigkeiten berücksichtigt werden und in das Training einbezogen werden. In dieser Dissertation werden Ansätze zur Identifikation der relevanten kognitiven Komponenten und zur Integration dieser Komponenten in den Trainingsprozess vorgestellt, um die Effizienz von traditionellem Training zu steigern und die Schwachstellen existierender Trainingssysteme zu verbessern.

Skill Acquisition and Training

Dieses Kapitel befasst sich mit den Basistheorien und -methoden der Aneignung und des Trainings von Fähigkeiten. Relevante Aspekte zur Entwicklung und Evaluierung von Programmen zum Training von Fähigkeiten werden zusammengetragen und beschrieben. Dies umfasst die Definition von Fähigkeiten *(skills)* und die Zerlegung von Fähigkeiten in Unterkategorien *(sub-skills)*, die Beschreibung von Modellen und Theorien zur Aneignung von Fähigkeiten (z.B. Fitts Drei-Stufen-Modell Theorie, das Dreyfus-Modell und Information Processing Theorien), die Erläuterung relevanter Trainingsmethoden und Aspekte/Methoden zum Messen und Bewerten von Fähigkeiten. Des Weiteren wird das Potential der Integration multimodaler Schnittstellen in den Aneignungs- und Trainingsprozess von Fähigkeiten herausgearbeitet und erörtert.

Es zeigt sich, dass für die Entwicklung effizienter Systeme zum Training von Fähigkeiten viele verschiedene Faktoren berücksichtigt werden müssen. So ist es zum Beispiel wichtig, die zu trainierende Fähigkeit in ihre zugrunde liegenden sensomotorischen und kognitiven Sub-Skills zu zerlegen, um zu identifizieren, welche dieser Sub-Skills im Training adressiert werden müssen. Darüber hinaus ist es wichtig den Lernprozess von Menschen zu verstehen und ihn durch Trainingsmethoden zu unterstützen, die sich bereits als gut erwiesen haben. Es wird auch deutlich, dass die Bereitstellung von Informationen mittels mehrerer Modalitäten (z.B. visuell, haptisch, akustisch) das Wahrnehmungsvermögen und die Auffassungskraft des Benutzers steigern kann, und somit den Trainingsprozess verbessern kann.

Technologies for AR-Based Skill Training

Ein weiterer wichtiger Faktor für die Entwicklung von Trainingssystemen ist der Einsatz geeigneter Technologien zur Umsetzung und Anwendung der entsprechenden Trainingsmethoden und zur Bereitstellung von Informationen und Feedback auf bestmögliche Art und Weise. Des Weiteren werden Technologien zum Erfassen von Informationen benötigt, um den Benutzer durch angemessenes Feedback zu unterstützen. In diesem Kapitel werden Technologien, die für das Training von Wartungs- und Montagefähigkeiten von Bedeutung sind, herausgearbeitet und diskutiert.

Eine dieser Technologien ist Augmented Reality (AR). In Augmented Reality werden reale Bilder mit virtuellen Objekten kombiniert. Computergenerierte, virtuelle Objekte werden mit Kamerabildern überlagert und in die Sicht des Benutzers eingeblendet. Die Realität wird mit zusätzlichen Informationen erweitert bzw. angereichert. Auf diese Weise können Informationen und Feedback gerichtet visualisiert werden.

Für die Bedien- und Benutzbarkeit eines Trainingssystems ist die Verwendung geeigneter Ein- und Ausgabegeräte ein wichtiger Faktor. Bei einem Augmented Reality-basierten System betrifft dies vor allem das verwendete Display Device. In diesem Zusammenhang werden verschiedene AR-Display Methoden (optical see-through und video see-through) und Devices (see-through devices, hand-held devices, head-attached displays und spatial displays) beschrieben und diskutiert. Daraus ergibt sich, dass sich vor allem hand-held Devices wie Smartphones und Tablet PCs als besonders geeignet zum Training von Wartungs- und Montagearbeiten erweisen. Aufgrund der hohen Rechenleistung heutiger Smartphones und Tablet PCs können komplexe Anwendungen direkt auf diesen Geräten ausgeführt werden, ein Auslagern von Berechnungsvorgängen auf externe (angebundene) Computer ist nicht länger notwendig. Des Weiteren verfügen diese Geräte zumeist über Schnittstellen zur Fingereingabe (touch interfaces), wodurch die Gestaltung einer intuitiven Bedienung von Anwendungen erheblich vereinfacht werden kann. Darüber hinaus bieten Smartphones und Tablet PCs eine höhere Auflösung als z.B. Head-Mounted Displays (HMDs), die von vielen Benutzern als unkomfortabel und unangenehm empfunden werden.

Um geeignetes Feedback bereitzustellen und um auf Aktionen des Benutzers während des Trainings zu reagieren, müssen verschiedene Arten von Informationen erfasst werden. Diese Faktoren umfassen unter anderem Bewegungen, ausgeübte Kräfte, zur Interaktion verwendete Objekte und kognitive Faktoren. Die Anzahl der Aspekte, die bei der Ausführung komplexer Fähigkeiten involviert sind, ist enorm, und es ist unmöglich alle dieser Aspekte zu Erfassen und während des Trainings zu berücksichtigen. Daher müssen die relevantesten Aspekte für die jeweiligen Trainingsaufgaben ausgewählt und mittels entsprechenden Technologien erfasst werden. Allerdings müssen Design und Architektur eines Trainingssystems die Integration vieler verschiedener Capturing-Technologien und das Verwalten und Verarbeiten der entsprechenden Capturing-Daten unterstützen, damit es zum Training unterschiedlicher Fähigkeiten eingesetzt werden kann (z.B. Training von Wartungs- und Montagefähigkeiten, Fähigkeiten zum Rudern oder Jonglieren, chirurgische Fähigkeiten, etc.). Die zum Training von Wartungs- und Montagearbeiten relevantesten Capturing-Technologien werden in diesem

Kapitel zusammengestellt und beschrieben. Dies beinhaltet das Tracking für Augmented Reality, Hand-Tracking, Motion Capturing und Force Capturing.

Um multimodales Feedback bereitzustellen, müssen Daten auf verschiedene Art und Weise bzw. durch verschiedene Modalitäten (visuell, haptisch und akustisch) präsentiert werden. Dies erfordert den Einsatz verschiedener Rendering-Technologien, die das Präsentieren von visuellen, haptischen und akustischen Informationen und Feedback ermöglichen, die ebenfalls in diesem Kapitel beschrieben werden.

AR-Based Multimodal Training

Für die Entwicklung eines Konzepts zum effizienten Training von Wartungs- und Montagefähigkeiten ist es wichtig zu untersuchen, welche Technologien eingesetzt werden können, um den Benutzer während des Trainings zu unterstützen, und wie sie angewendet werden sollten, um ihm die bestmögliche Unterstützung zu bieten. Ein weiterer kritischer Faktor ist, wie Informationen und Feedback während des Trainings präsentiert werden. In diesem Kapitel wird der Einsatz von multimodalem Feedback und Augmented Reality Technologien in Trainingsumgebungen herausgearbeitet und diskutiert. Dabei werden die Verwendung von vibrotaktilem Feedback als zusätzlicher Informationskanal in AR-basierten Trainingsumgebungen und die Visualisierung von Informationen in Augmented Reality-basiertem Training im Detail analysiert. Des Weiteren wird ein neues Visualisierungskonzept für die Präsentation von Informationen in Augmented Reality Trainingsumgebungen eingeführt, welches Probleme existierender AR-Trainingssysteme kompensiert und durch die Integration verschiedener Medientypen (z.B. Text, Bilder, Videos, 3D-Modelle und -Animationen) sowohl ein enormes Potential an Informationsvermittlung bietet, als auch die Generierung der Inhalte für AR-basierte Trainingsanwendungen erheblich vereinfacht, da vorhandenes (bewährtes) Trainingsmaterial leicht integriert werden kann. Schließlich wird auch ein Designkonzept einer multimodalen Augmented Reality-basierten Trainingsplattform vorgestellt und beschrieben.

Bei der Analyse des Einsatzes von multimodalem Feedback während des Trainings zeigt sich, dass die Integration verschiedener Modalitäten in ein Trainingssystem das Training erheblich verbessern kann. Da das Sehen in den meisten Fällen die primäre Modalität darstellt, ist die Verfügbarkeit visueller Informationen und deren qualifizierte Präsentation bedeutende Faktoren für die Entwicklung von effizienten Trainingssystemen. Führung und Instruktion sind wichtige Aspekte des Trainings, sie dürfen jedoch nicht des Benutzers aktive Erkundung der Trainingsaufgabe behindern. Daher dürfen die verschiedenen Modalitäten nicht kombiniert werden, um die Führung des Benutzers zu verstärken, sondern vielmehr um lenkende ("channeling") Informationen zu generieren, die den Benutzer in der Erkundung der Aufgabe unterstützen und ihn leicht zu dem angestrebten Ziel leiten. Eine wichtige Erkenntnis ist, dass die Präsentation multimodaler Informationen die Performanz des Benutzers steigern kann. Durch eine Übertragung dieser Erkenntnis in Augmented Reality-Umgebungen zeigt sich, dass eine multimodale Augmentierung der Realität zur Unterstützung des Benutzers während der Ausführung des Trainingstasks eingesetzt werden kann und soll.

Aufbau und Struktur von AR-basiertem Training sind ebenfalls bedeutende Aspekte. Daher legt dieses Kapitel auch Augenmerk auf die Entwicklung eines Trainingsprogramms zum Training in multimodalen AR-Umgebungen. In diesem Zusammenhang werden sogenannte Accelerators eingeführt. Diese stellen Variablen dar, die in den Trainingsprozess eingeführt werden, um das Training zu erleichtern, aufzuwerten und zu verbessern. Dieses Konzept der Accelerators wird eingeführt, um zu definieren, wie Technologien zur Verbesserung des Trainings angewendet werden können. Zu den Hauptaufgaben der Accelerators gehört es, den Benutzer dabei zu führen seine Befähigung bestmöglich zu nutzen. Des Weiteren sollen sie die Performanz des Benutzers in der Ausführung der Trainingsaufgabe verbessern und somit die Trainingseffizienz steigern. Im Kontext von Augmented Reality-basiertem Training bedeutet dies, dass Accelerators definieren, wie eine oder mehrere verfügbare Technologien eingesetzt werden können um (1) die Lern-/Trainingszeit zu reduzieren und/oder (2) die Performanz des Benutzers in der Ausführung der Fähigkeit zu verbessern.

Wie bereits erwähnt, wird darüber hinaus ein neues Konzept zur Visualisierung von ortsabhängigen Informationen in Augmented Reality-Umgebungen eingeführt, durch welches Trackingungenauigkeiten kompensiert werden können. In diesem Konzept wird die zum Annotieren von Dokumenten typische *Zeiger-Inhalt* Metapher in Augmented Reality-Umgebungen übertragen. Als Ergebnis werden Adaptive Visual Aids (AVAs) definiert, die aus einem trackingabhängigen Zeigerobjekt und einem trackingunabhängigen Inhaltsobjekt bestehen. Informationsgehalt und Art der Information (Text, Bild, Video, 3DInhalte) von Zeiger- und Inhaltsobjekt sind regulierbar. Auf diese Weise kann die Stärke der Führung des Benutzers, die durch überlagerte 3D-Objekte in Augmented Reality bereitgestellt wird, einfach gesteuert werden. Somit kann verhindert werden, dass der Benutzer in der aktiven Erkundung der Trainingsaufgabe gestört wird. Adaptive Visual Aids können auch benutzt werden, um traditionelle Augmented Reality-Einblendungen (d.h. 3D-Modelle und -Animationen), die sehr unter Trackingungenauigkeiten leiden, zu ersetzen. Außerdem können vorhandene Trainingsmaterialen, wie z.B. Bilder, Videos oder Beschreibungen in Form von Text, im Inhaltsobjekt einer Adaptive Visual Aid eingebunden werden.

Zur Entwicklung eines Trainingssystems müssen zunächst die Hauptbestandteile eines solchen Systems erarbeitet und definiert werden. Für ein multimodales AR-basiertes Trainingssystem lassen sich folgende drei Hauptkomponenten definieren: die physikalische Trainingsumgebung, die multimodale Trainingsplattform und ein digitales Daten-Repository. Die physikalische Trainingsumgebung umfasst alle physikalischen Objekte, die in das Training bzw. in die Trainingsaufgabe involviert sind (z.B. die zu wartende Maschine, Werkzeuge, etc.), ebenso wie die realen Umgebungsgegebenheiten, wie zum Beispiel Beleuchtung und Temperatur. Des Weiteren beinhaltet sie alle Schnittstellen, die an der Verbindung der realen Welt mit dem Computer/den Computern beteiligt sind, wie z.B. Display Devices und Kameras. Das digitale Daten-Repository ist eine Art Datenbank, die Daten wie Workflow-Beschreibungen der Aufgabe, Trainingsprotokolle, Trainingsmaterial (z.B. Visual Aids Objekte, Sprach-instruktionen, Bilder, Videos), benutzerspezifische Daten (z.B. aktuelle Trainingsphase), oder auch bereits erfasste Capturing-Daten enthält. Die Hauptaufgaben der multimodalen Trainingsplattform sind das Erfassen von Daten (Capturing), die Auswahl angemessener Instruktionen, die Generierung von entsprechendem Feedback und die Präsentation von Informationen und Feedback für den

Benutzer. Die Plattform verwaltet die komplette Logik der Trainingsanwendung, wie zum Beispiel die Entscheidung des nächsten Schrittes in der Trainingsaufgabe und die Auswahl entsprechender Instruktionen und Feedback unter Berücksichtigung der aktuellen Trainingsphase des Benutzers. Die Plattform enthält unter anderem auch Machine Learning-Komponenten (z.B. Hidden Markov Model-Mechanismen), die zur Auswahl von geeignetem Feedback und zur Selektion des nächsten Schrittes verwendet werden können. Diese Aufgaben der multimodalen Trainingsplattform werden mittels drei Bausteinen der Plattform bewerkstelligt. Diese Bausteine—multimodaler Capturing Controller, Interaction Processing and Application Module (IPA) und multimodaler Rendering Controller—werden in der Arbeit im Detail vorgestellt.

Das Designkonzept der multimodalen Trainingsplattform, das in dieser Arbeit vorgestellt wird, ist nicht spezifisch für das Training von Wartungs- und Montagefähigkeiten entwickelt, sondern stellt vielmehr ein generelles Designkonzept für multimodale Trainingsplattformen dar. Darüber hinaus wird in diesem Konzept nicht nur das Training von Benutzern berücksichtigt, sondern auch die Idee des Trainieren des Systems bzw. der Trainingsanwendung durch Experten. Dies ist ebenfalls ein bedeutsamer Aspekt, da durch das Erfassen des Verhaltens von Experten existierende Workflow-Beschreibungen der Trainingsaufgabe verfeinert oder gar neu generiert werden können, um so eine bestmögliche Beschreibung der Fähigkeit bzw. des der Fähigkeit zugrunde liegenden Workflows zu erhalten.

AR-Based Training of Maintenance and Assembly Skills

Für die Entwicklung eines effizienten Trainingskonzepts für das Training von Wartungs- und Montagefähigkeiten ist die Identifikation der zugrunde liegenden Sub-Skills, die für die Ausführung und Aneignung von Wartungs- und Montageprozeduren relevant sind, von enormer Bedeutung. Darüber hinaus muss erarbeitet werden, wie diese Sub-Skills effizient trainiert werden können. Dieses Kapitel adressiert sowohl die Identifikation dieser Sub-Skills, also auch das Training von Wartungs- und Montagefähigkeiten in multimodalen Augmented Reality-Umgebungen unter Einbeziehung der relevantesten Sub-Skills. Die Analyse von Wartungs- und Montagefähigkeiten zeigt, dass die sogenannten prozeduralen Fähigkeiten (Procedural Skills), als relevanteste zugrunde liegenden Sub-Skills betrachtet werden können. Daher wird besonderes Augenmerk auf Elemente zum Training dieser prozeduralen Fähigkeiten gelegt. Unter prozeduralen Fähigkeiten versteht man die Fähigkeit, wiederholt einer Reihe von Aktionen Schritt für Schritt zu folgen, um ein bestimmtes Ziel zu erreichen. Durch Recherche vorhandener Literatur und in Kooperation mit Kognitionswissenschaftlern wurden die wichtigsten Aspekte zum Training von prozeduralen Fähigkeiten herausgearbeitet und werden in der Arbeit beschrieben. Ein bedeutender Aspekt stellt das so genannte mentale Modell des Benutzers dar (d.h. des Benutzers interne Repräsentation der Aufgabe). Der Aufbau dieses mentalen Modells kann durch die Einbeziehung unterschiedlicher Modalitäten in das Training und durch die Bereitstellung von Kontext- und Strukturinformationen unterstützt werden.

Basierend auf dieser Erkenntnis wird in der Dissertation ein Trainingsprogramm für multimodales AR-basiertes Training von Wartungs- und Montagefähigkeiten vorgestellt. Dies umfasst die Definition spezifischer Trainingsstrategien und Accelerators, sowie die Entwicklung eines Trainingsprotokolls. Die verfolgten Trainingsstrategien bestehen aus der Bereitstellung von direkten Hilfselementen (direct aids), indirekten Hilfselementen (indirect aids) und kontextbezogenen Hilfsmitteln (context aids). Diese Strategien werden durch folgende Accelerator realisiert:

- *Adaptive Visual Aids mit Information auf Nachfrage*: Durch die Regulierung des Führungslevels, was durch die Verwendung von AVAs ermöglicht wird, wird der Benutzer nicht zu stark geführt, sondern vielmehr gefordert, und nicht in der aktiven Erkundung der Aufgabe behindert. Ebenso wird er im Aufbau eines mentalen Modells der Aufgabe unterstützt.
- *Device Display*: Das Device Display ist ein visuelles Element, das Informationen über einzelne Unteraufgaben (sub-tasks) einer Aufgabe, bzw. über Schritte, die zu einer logischen Einheit gehören. Das heißt, es visualisiert ein gutes mentales Modell und unterstützt somit den Benutzer in der Erstellung einer internen Repräsentation der Aufgabe.
- *Struktur- und Kontextinformationen*: Ebenso die Bereitstellung von Struktur- und Kontextinformationen kann den Aufbauprozess eines mentalen Modells unterstützen, da auf diese Weise dem Benutzer zusätzliche Informationen über die Struktur der Aufgabe vermittelt werden können. Als Beispiel für ein solches Informationselement wird der sogenannte erweiterte Fortschrittsbalken eingeführt, der sowohl Informationen über den Fortschritt des Benutzers in der kompletten Aufgabe, als auch Informationen über die einzelnen Sub-Tasks, bzw. den Fortschritt des Benutzers innerhalb eines Sub-Tasks, bietet.

Die Entwicklung des Trainingsprotokolls beinhaltet die Definition von Aufgaben-Szenarien (bzw. Trainings-Szenarien), die Beschreibung des Trainingsprozesses, sowie die Definition von messbaren Leistungsindikatoren des Benutzers. Diese Indikatoren bestehen aus:

- *Ausführungszeit der Aufgabe*: Die Zeit, die der Benutzer zur Bewerkstelligung der Trainingsaufgabe benötigt.
- *Ausführungszeit der Schritte*: Die Zeit, die der Benutzer zur Bewerkstelligung der einzelnen Schritte der Aufgabe benötigt.
- *Anforderungen von Hilfselementen*: Die Anzahl der vom Benutzer veranlassten Aktivierungen der nicht permanent visualisierten Informationen (z.B. Inhaltsobjekte von Adaptive Visual Aids).

Durch die Auswertung dieser Leistungsindikatoren kann die Performanz des Benutzers bewertet werden.

Setup and Implementation of the Training Platform

Das Hardware Setup der entwickelten Trainingsplattform besteht aus vier Hauptkomponenten:

- Tablet PC
- Webcam, die an den Tablet PC aufgesteckt wird
- Haptisches bzw. vibrotaktiles Armband

- Speziell entworfener, beweglicher Befestigungsarm zum Ablegen des Tablet PCs

Basierend auf den Erkenntnissen, die aus der Analyse verschiedener Augmented Reality Display Devices resultieren, und auf den aufgezeigten Vorteilen von hand-held Geräten wird für die Umsetzung des Systems zur Evaluierung der Plattform ein Tablet PC mit Fingereingabe-Unterstützung verwendet. Die verwendete Webcam dient zur Aufnahme der realen Bilder für die Berechnung der Augmented Reality-Überlagerungen. Da noch kein Standard für Augmented Reality existiert, wurde das entwickelte Plattformkonzept auf Basis des X3D ISO Standards implementiert. Dieser Standard wurde ausgewählt, da er Merkmale bereitstellt, die für die Entwicklung von Augmented Reality-Umgebungen nützlich sind. So ist zum Beispiel durch die Verwendung des X3D Standard für die Implementierung eine saubere Trennung zwischen Applikation und System erreicht worden. Somit kann die Trainingsapplikation an neue Szenarien angepasst werden, ohne das zugrunde liegenden System modifizieren zu müssen. Des Weiteren ist es dadurch auch Nicht-Programmierern möglich, die Trainingsanwendung zu ändern oder zu erweitern. Die Implementierung der Trainingsplattform, der Trainingsanwendung und der Accelerators wird in der Dissertation im Detail beschrieben.

Eine solch standardbasierte Implementierung ist ebenso ein wichtiger Faktor zur Steigerung der Nachhaltigkeit und Portabilität der Plattform. Durch die Bereitstellung wohldefinierter Schnittstellen wird die Anbindung neuer Geräte und Bibliotheken erheblich vereinfacht, wodurch die Akzeptanz der Trainingsplattform enorm gesteigert wird.

Evaluation of the Training Platform

Die implementierte Plattform und die oben beschriebenen Accelerators wurden in der Industrie evaluiert und den dort aktuell angewendeten traditionellen Trainingsmethoden gegenübergestellt. Sowohl Trainer als auch Techniker waren an der Evaluierung beteiligt. Die Evaluierung wurde in zwei Stufen durchgeführt: Als erste Stufe eine *Usability & Functionality* Evaluierung, als zweite Stufe eine *Skill Transfer* Evaluierung.

In der *Usability & Functionality* Evaluierung dienten vier Trainer als Teilnehmer. Hier basierte die Evaluierung der Plattform und der Accelerators auf subjektiven Einschätzungen der Trainer. Das heißt, nach der Verwendung der Plattform zum Training des definierten Task Szenarios füllten die Trainer Fragebögen über die implementierten Funktionalitäten und die Bedienbarkeit der Plattform aus. Dies beinhaltete unter anderem Fragen und Bewertungen bezüglich der bereitgestellten Informationen (z.B. Instruktionen), der verwendeten Hilfsmittel (z.B. vibrotaktile Informationen), des Aufbaus der Trainingsapplikation und der Bedienbarkeit der Trainingsapplikation. Durch Auswertung dieser Evaluierungsstufe konnte die Plattform in ihrer Bedien- und Benutzbarkeit bewertet und gegebenenfalls angepasst werden.

Die zweite Evaluierungsstufe, die *Skill Transfer* Evaluierung, wurde anhand 20 Techniker durchgeführt. Diese wurden in zwei Gruppen mit jeweils zehn Technikern aufgeteilt. Eine Gruppe trainierte die definierte Trainingsaufgabe unter Verwendung der implementierten Plattform (AR-Gruppe), die andere Gruppe trainierte diese Aufgabe mittels Standard-Trainingsmethoden (Kontrollgruppe). Besonderen Wert wurde auf die Homogenität der Gruppen gelegt. Um diese Homogenität zu überprüfen, mussten alle Teilnehmer vorab demographische Fragebögen ausfüllen. Des Weiteren wurden sogenannte Fähigkeitstests (capability tests) durchgeführt, um das aktuelle Fähigkeitslevel der Teilnehmer festzustellen. Zur Evaluierung der entwickelten Plattform führten alle Teilnehmer die Trainingsaufgabe an einem späteren Tag ohne Verwendung von Hilfsmitteln durch. Dabei wurden verschiedene Faktoren gemessen und ausgewertet, wie zum Beispiel:

- Ausführungszeit: Die Ausführungszeit der Aufgabe
- Gelöste Fehler: Die Anzahl der Fehler, die der Teilnehmer zwar gemacht hat, aber während der Ausführung der Aufgabe ohne fremde Hilfsmittel korrigieren konnte
- Nicht-gelöste Fehler: Die Anzahl der Fehler, die der Teilnehmer gemacht hat, aber nicht korrigieren konnte
- Anforderungen von Hilfsmitteln: Die Anzahl der Anforderungen von fremden Hilfsmitteln während der Ausführung der Aufgabe

Darüber hinaus wurde in Kooperation mit den Trainern der Industrie und Kognitionswissenschaftlern ein *Scoring* definiert, anhand dessen der Fähigkeitslevel der Teilnehmer unter Verwendung der gemessenen Leistungsindikatoren bestimmt wurde. Das Ergebnis der Skill Transfer-Evaluierung zeigt, dass die entwickelte Trainingsplattform sehr gut zum Training von Wartungs- und Montagefähigkeiten geeignet ist, und dass dadurch traditionelles Training verbessert werden kann. Die Ausführungszeit der Trainingsaufgabe war für die Teilnehmer der AR-Gruppe im Mittel zwar nur wenig kürzer, allerdings machte die AR-Gruppe deutlich weniger nicht-gelöste Fehler. Des Weiteren konnte auch der Fähigkeitslevel der Teilnehmer der AR-Gruppe im Vergleich zur Kontrollgruppe als höher bewertet werden.

Somit wird deutlich, dass mit dem vorgestellten Framework die oben skizzierten Probleme bewältigt werden. Aufwand und Kosten für das Training von Wartungs- und Montagefähigkeiten werden reduziert, die Trainingseffizienz im Vergleich zu traditionellem Training wird gesteigert.

Conclusions

Diese Arbeit stellt ein neues Konzept für die Gestaltung von effizientem Training von Wartungs- und Montageprozessen unter Verwendung von Augmented Reality-Technologien vor, das sowohl Visualisierungsparadigmen zum Präsentieren von Instruktionen und Informationen in multimodalen Augmented Reality-Trainingsumgebungen, als auch ein Designkonzept für die Entwicklung einer multimodalen Trainingsplattform zur Umsetzung des Trainings beinhaltet.

Es wird ein neues Konzept zur Visualisierung von ortsabhängigen Informationen in Augmented Reality-Umgebungen eingeführt, welches Probleme bestehender Systeme behebt und die Anwendbarkeit von Augmented Reality in industriellen Umgebungen steigert. In diesem Konzept wird die zum Annotieren von Dokumenten typische *Zeiger-Inhalt* Metapher in Augmented Reality-Umgebungen übertragen. Als Ergebnis werden *Adaptive Visual Aids (AVAs)* definiert, die aus einem trackingabhängigen Zeigerobjekt und einem trackingunabhängigen Inhaltsobjekt bestehen. Informationsgehalt und Art der Information (Text, Bild, Video, 3D-Inhalte) von Zeiger- und Inhaltsobjekt sind regulierbar. Auf diese Weise kann die Stärke der Führung des Benutzers, die durch überlagerte 3D-Objekte in Augmented Reality bereitgestellt wird, einfach gesteuert werden. AVAs können auch benutzt werden, um traditionelle Augmented Reality-Einblendungen (d.h. 3D-Modelle und -Animationen), die sehr unter Trackingungenauigkeiten leiden, zu ersetzen. Außerdem können vorhandene Trainingsmaterialen, wie z.B. Bilder, Videos oder Beschreibungen in Form von Text, im Inhaltsobjekt einer AVA eingebunden werden, wodurch der Erstellungsprozess der Inhalte von Augmented Reality-Anwendungen im Allgemeinen und von AR-basierten Trainingsanwendungen im Speziellen wesentlich beschleunigt werden kann.

Das Designkonzept der multimodalen Trainingsplattform, das in dieser Arbeit vorgestellt wird, ist nicht spezifisch für das Training von Wartungs- und Montagefähigkeiten entwickelt, sondern stellt vielmehr ein generelles Designkonzept für multimodale Trainingsplattformen dar. Dieses Konzept berücksichtigt auch die Möglichkeit, Trainingsprotokolle durch Experten zu trainieren, um zum Beispiel die Modellierung des entsprechenden Arbeitsflusses des Trainingsszenarios (Workflow) zu optimieren. Des Weiteren wurde gezeigt, wie das entwickelte Plattformkonzept auf Basis des X3D ISO Standards implementiert werden kann.

Contents

1. Introduction

1.1. Motivation

As the complexity of maintenance and assembly tasks can be enormous, the training of technicians to acquire the necessary skills to perform these tasks efficiently is a challenging point. Good guidance of the user through the training task is one of the key features to improve the efficiency of training. Traditional training programs are often expensive in terms of effort and cost, and rather inefficient, since the training is highly theoretical. Usually, the training sessions take place in special classrooms (training centers) equipped with state-of-the-art audio-video tools for remote conferencing, teaching and e-learning. With training material (e.g. slides) presented by a trainer, the technicians are taught how to assemble, disassemble and repair the machine parts. If available, practical training, in which the technicians learn the practical execution of the maintenance tasks, is conducted in separate training sessions in dedicated labs. Usually, a trainer is on-site or at least remotely connected, to give instructions and hints.

The preparation of the required training material is a critical factor in the development of a training program. The authoring of meaningful slides for the presentations and explanatory documents such as hand-outs requires a lot of time and effort, and is usually specific for one dedicated task. Hence, specific training material needs to be prepared for each task. Furthermore, the traditional training of technicians is an expensive and drawn-out process that requires a lot of training cycles for each specific task.

As already mentioned, the complexity of maintenance tasks is increasing. Thus, it is not sufficient to train the technicians' execution of maintenance tasks, but rather to train the underlying skills. That is, the skills—sensorimotor and cognitive—required for the efficient acquisition and performance of new tasks have to be trained as well. In this context, the acquisition, interpretation, storage and/or simulation of sensorimotor and cognitive skills are becoming increasingly important. This leads to three major demands that skill-training systems should meet: speed, efficiency, and transferability of training. Therefore, training procedures need to be developed which facilitate or rather accelerate the trainee's learning process. To be more precise, these training procedures should support the trainee in the process of observing a model, establishing adequate representations of motor and behavioral responses, transforming specific instructions into appropriate actions, extending simple sensorimotor and behavior schemes to new and more complex conditions, and transferring skills from one situation to another.

Previous research has shown that Augmented Reality (AR) is a powerful technology to support training in the particular context of industrial service procedures. Instructions on how to assemble/disassemble a machine can be directly linked to the machines to be operated. Various approaches exist, in which the trainee is guided step-by-step through the maintenance task. Mostly technical aspects (tracking, visualization etc.) have been the focus of this research field. Furthermore, those systems act more as guiding systems than as training systems and focus only on the trainees' sensorimotor capabilities.

A potential danger of Augmented Reality applications is that users become dependent on Augmented Reality features, such as visual instructions, for the performance of a task. As a result, the user might not be able to perform the task when those features are not available or when the technology fails. That is, an Augmented Reality-based training system must clearly differ from an Augmented Reality-based guiding system; it must really train the user and challenge him to actively explore the task instead of simply guiding him through the training task to the designated goal. This can only be achieved by involving cognitive aspects in the training.

All these facts lead to the need for efficient training systems which accelerate the technicians' learning and acquisition of new maintenance procedures. Furthermore, these systems should improve the adjustment of the training process to new training scenarios and enable the reuse of existing training material that has proven its worth.

1.2. Challenges

One of the big challenges in current research on efficient training systems is the *combination of sensorimotor and cognitive skills training*. If the training of sensorimotor and cognitive capabilities is completely separated then it is difficult for the trainee to establish a relationship between both aspects since the training of cognitive skills is usually highly abstract. Hence, the trainee can not directly apply the acquired "abstract" knowledge to the concrete task to train. Therefore, the training of both sensorimotor and cognitive skills has to be combined in one training system that can be applied to the real training task (e.g. disassembling a car engine). Another important factor is the identification of the sensorimotor and cognitive sub-skills of the complex skill (e.g. maintenance skill) to be trained.

A further challenge is the design of the *multimodal interface*. Various feedback channels (visual, haptic, audio, etc.) have to be integrated and controlled to exploit the potential of multimodality for training purposes (see Chapter 3.5). The concept must also cope with multimodal capturing technologies, such as computer vision, EMG and force sensors, since the workflow ("which step is next?") and the corresponding feedback provided to the trainee depends on captured information. The multimodal interface must be designed in such a way that the system can be adapted to different training scenarios with minimum effort. In addition, the system should be scalable to run on mobile devices to allow training in the real environment under real conditions and avoid the need for dedicated training labs. To make the system portable and, even more important, sustainable, the concept should allow for a standards-based implementation.

Adaptability and robustness of the system are also crucial factors for the development of training systems. In order to provide a wide range of different training scenarios the system must be easily adaptable to different training tasks and conditions. The ability to change training conditions is also important to support different training phases and different skill levels of the trainees (e.g. intermediate learner, advanced learner). To reduce the effort in creating new training scenarios (for new tasks) the adaption of corresponding system parameters and feeding the system with new content should require minimum effort. Furthermore, it should be possible to integrate existing, reliable training material. A good adaptability of a training system to different training tasks and conditions is essential for its acceptance by the training institution (e.g. the industry), as well as the robustness of the system. The system should work robustly in real working environments (e.g. a machinery hall).

A final crucial factor is the *generalizability* of the concept. The applicability of a concept to various application fields is important for its sustainability and impact. Hence, the concept should be transferable to other application domains, such as rehabilitation or sport.

1.3. Contributions and Overview

The main goal of this thesis is to provide an efficient multimodal training platform for improving technicians' maintenance and assembly skills in order to accelerate their acquisition process of new maintenance and assembly tasks. A concept for multimodal Augmented Reality based training of maintenance skills has been developed which fulfils the requirements of efficient training. The work presents results of interdisciplinary research based on the fusion of cognitive science and computer science in the field of maintenance and assembly training. To overcome limitations of existing approaches, the work addresses the combined training of both sensorimotor and cognitive skills by exploiting the potential of Augmented Reality technologies and multimodal interfaces. Novel concepts and techniques for visual augmentation have been developed in order to improve the visual feedback for the user and to cope with problems caused by tracking inaccuracies. This work also provides key elements—guidelines—of maintenance skill training, called *Accelerators*, which can be transferred to any procedural training task. It is shown how such a multimodal Augmented Reality environment providing on-the-job training can be implemented based on a standard for real-time 3D communication (X3D [Web08a]). In the focus of this work is the presentation component of the training system, that is, the multimodal presentation of feedback to the trainee, whereby amount and type of feedback can depend on captured input data.

The developed concept includes the possibility to easily adapt the training system to different training conditions in order to support various phases of a training program and trainees at various skill levels. Furthermore, the system is able to reuse existing reliable training material. In addition, the generation of new training task scenarios is simplified in comparison to existing training systems.

Due to all these factors, the developed system improves the training process of industrial maintenance tasks. The training of technicians in acquiring new maintenance tasks, a critical cost factor in industry, is accelerated and, hence, training cycles can be abbreviated.

Chapter 2 provides an overview over existing approaches of training systems based on Augmented Reality or multimodal interfaces in the fields of industry and manufacturing, surgery and rehabilitation, and sport and entertainment.

In Chapter 3 basic theories and methodologies of the acquisition and training of skills are described. This includes amongst others the definition of the term *skill* and a concept for the decomposition of complex skills into sub-skills. These sub-skills comprise sensorimotor skills as well as cognitive skills. Basics on information processing (open-loop, closed-loop theories) and two different models of skill acquisition are presented and briefly described. The Three-Stage Model of skill acquisition developed by Fitts [Fit54] states, that the development of skills proceeds through three stages: a cognitive stage, an associative stage and an autonomous stage. In contrast, the Dreyfus-Model [DD80] indicates, that a learner passes through five levels of proficiency when acquiring a skill: novice, advanced beginner, competent, proficient, and expert. Furthermore, some training methodologies and aspects which are relevant for this work are explained. Subsequently it is shown how skills can be measured. Finally, the potential of multimodality for training is discussed.

Chapter 4 explains basic technologies for the training of skills. Firstly, an introduction to Augmented Reality and corresponding display devices is given. Secondly, capturing and rendering technologies that are relevant for skill-training systems are described. The technologies presented include force capturing, camera tracking, hand tracking, and motion capturing, as well as visual, haptic, and audio rendering.

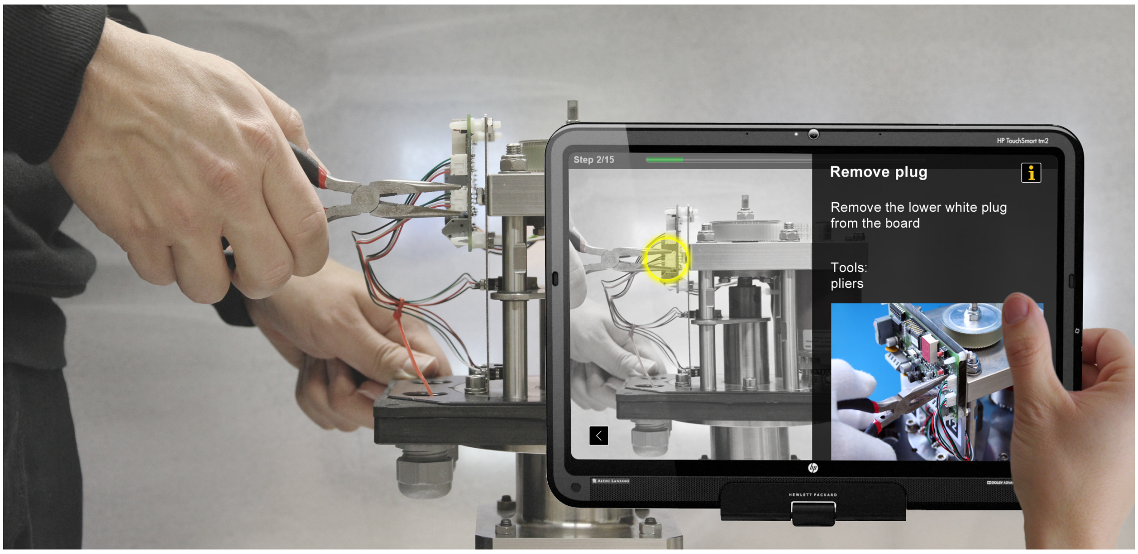


Figure 1.1.: Visual augmentations in the developed Augmented Reality-based training system for maintenance and assembly skill training.

Chapter 5 addresses the training of skills in multimodal Augmented Reality environments. The significance of multimodal training environments and the use of different modalities for providing feedback are discussed, and the establishment of such a training environment is described. In addition, the concept of using *accelerators* is introduced. Accelerators are variables which are integrated in the training process and implemented to facilitate, assist, and improve learning. This concept is not specific to maintenance and assembly processes and can be transferred to other application domains. In this chapter, the potential of vibrotactile feedback in multimodal training environments is also discussed, and the results of evaluation studies are presented. Furthermore, a novel visualization concept for Augmented Reality-based instructions and training-oriented information is provided. This concept addresses the type and amount of information provided by visual augmentations. For example, a visual augmentation can consist of a superimposed, animated 3D model showing the task to perform (the traditional way), but also of an annotation that consists of a spatially aligned *pointer* highlighting the part of interest and screen-aligned *content* providing the detailed multimedia-based information on user demand (see Figure 1.1). At the end of the chapter, a design approach for AR-based multimodal skill training systems is introduced. The main components of such training systems are described and an overview over the architecture of the underlying training platform is given.

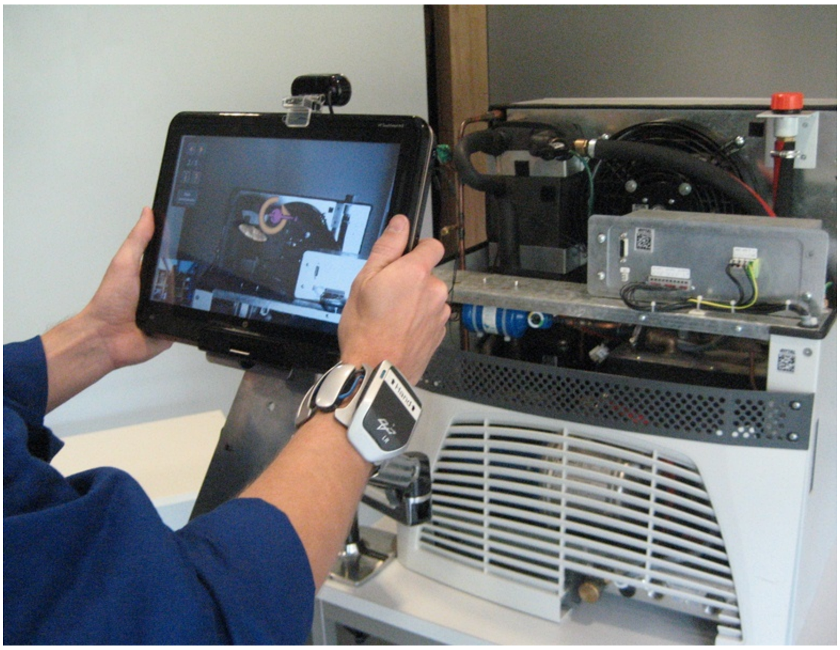


Figure 1.2.: The developed Augmented Reality based training system includes a vibrotactile bracelet for providing additional feedback.

A concept for Augmented Reality-based training of maintenance and assembly skills using multi-modal interfaces is presented in Chapter 6. First, the identification of the sensorimotor and cognitive sub-skills which are necessary to acquire maintenance skills (i.e. skills that are required to efficiently acquire and perform maintenance tasks) is discussed. Concepts for training these specific sub-skills have been developed and are presented in this chapter. The so-called *procedural skill*, which is the ability to repeatedly follow a set of actions step-by-step in order to achieve a specified goal, is identified as the most relevant sub-skill for the acquisition of maintenance skills. Elements for the training of procedural skills are analyzed in detail. Furthermore, specific accelerators for maintenance skills training that have been examined are presented in this chapter. These accelerators are (1) the use of visual aids with adjustable information level, (2) the visualization of a device display, (3) the provision of information about the structure of the training task and the user's progress in the task performance, and (4) the integration of additional feedback channels such as haptic hints. Also, the development of a training protocol for the training of an exemplary maintenance task is presented. This includes approaches for the evaluation of the specific accelerators and a training program for the integrated maintenance training system.

In Chapter 7 the realization of the multimodal Augmented Reality-based training platform is explained (see Figure 1.2). The importance of basing the realization of the platform on an established standard is pointed out, and it is shown how the platform can be implemented by exploiting and extending the *X3D* standard. In addition, the setup of the training system and the implementation of the defined accelerators and of the training application within an X3D environment are illustrated.

Finally, the evaluation of the training system conducted at the food packaging manufacturer *Sidel* is described in Chapter 8. Structure and performance of the evaluation are described and results are presented and analyzed.

2. Augmented Reality-Based and Multimodal Training Systems

Research on the design and development of training systems has increased significantly over the last years. This research covers various fields, such as industry, surgery, rehabilitation, and sports. There are many approaches coping with the challenges of integrating multimodal interfaces and advanced training environments, such as Augmented Reality environments. In this chapter some of these approaches in the different application areas and interesting studies on the design of training systems and corresponding interfaces are described.

2.1. Industry and Manufacturing

As the complexity of maintenance and assembly procedures can be enormous, the training of operators to perform those tasks efficiently has been in focus of many research groups. Numerous studies presented the potential of Augmented Reality based training systems and its use in guidance applications for maintenance tasks. One of the first approaches is using Augmented Reality for a photocopier maintenance task [FMS93]. The visualization is realized using wireframe graphics and a monochrome monoscopic Head Mounted Display (HMD). The tracking of objects and the user's head is provided by ultrasonic trackers. The main objective is to extend an existing two dimensional automated instruction generation system to an augmented environment. Hence, only simple graphics are superimposed instead of complicated 3D models and animations.

Schwald et al. [SLSG03] describe an AR system for training and assistance in the industrial maintenance context, which guides the user step-by-step through training and maintenance tasks. Magnetic and infrared optical tracking techniques are combined to obtain a fast evaluation of the user's position in the whole set-up and a correct projection for the overlays of virtual information in the user's view. The user is equipped with a lightweight helmet, which integrates an optical see-through HMD, a microphone, headphones, and a 3D-positioning sensor. The headphones offer the user the possibility to get audio information on the procedures to achieve. Via the microphone the user can easily interact with the system by using speech recognition. The 3D-positioning sensor is used to determine the position of the objects of interest in 3D space in relation to the user's position. That way, 3D augmentations are directly superimposed with their real counterparts, whereby the parts of interest are highlighted. Besides, also information about how to interact with the counterparts can be visualized. The paper discusses the usage of the system, the user equipment, the tracking and the display of virtual information.

Reiners et al. [RSKM98] introduce an Augmented Reality demonstrator for training a doorlock assembly task. The system uses CAD data directly taken from the construction/production database as well as 3D-animation and instruction data prepared within a Virtual Prototyping planning session, to facilitate the integration of the system into existing infrastructures. For the tracking they designed an optical tracking system using low cost passive markers. A HMD acts as display device.

In [OGL08] a spatial AR system for industrial Computerized Numerical Control(CNC) machines, that provides real-time 3D visual feedback by using a transparent holographic element instead of using user worn equipment (like e.g. a HMD). Thus, the system can simultaneously provide bright imagery and clear visibility of the tool and workpiece. To improve the user's understanding of the machine operations, visualizations from process data are overlaid over the tools and workpieces, while the user can still see the real machinery in the workspace. Also information on occluded tools is provided. The system, consisting of software and hardware, requires minimal modifications to the existing machine. The projectors need only to be calibrated once in a manual calibration process.

An Augmented Reality application for training and assisting in maintaining equipment is presented in [KKCL05]. Overlaid textual annotations, frames and pointing arrows provide information about machine parts of interest. That way, the user's understanding of the basic structure of the maintenance task and object is improved. A key component of the system is a binocular video see-through HMD, that the user is wearing. The tracking of the position and orientation of equipment is implemented using the ARToolKit [KB99].

Weidenhausen et al. [WKD03] developed and evaluated an Augmented Reality system for industrial applications by implementing and analyzing different scenarios. The tracking of the user's head, interaction and authoring is considered in this work. Most of the implemented scenarios applied a video-based tracking approach using markers, in which a camera mounted near the display (head-mounted- or hand-held-display) catches the scene. Furthermore, the authors pointed out, that users judge the usefulness of a technology mainly by the user interface.

Tümler et al. [TDM*08] focus in their work on the question, how long term usage of Augmented Reality technology in industrial training and assisting applications produces stress and strain for the user. They present user studies comparing strain during an AR supported and a non-AR supported work task. The strain is measured by analyzing the heart rate variability. Optimal Augmented Reality systems (no lag, high quality see-through calibration, weight-reduced HMD etc.) are considered as well as non-optimal ones. They showed that the use of optimal AR systems may decrease the overall strain, so that working with the AR system could mean a decreased strain compared to traditional work assistance systems.

An Augmented Reality system for fire support team training is presented in [BBb*05]. The system provides a HMD for the trainee and a touch screen for the instructor, who has to determine the effect on the target. Both HMD and touch screen are showing virtual moving targets on the real range. The trainee, whose task is to detect a target and call for fire, can see a simulated magnified view and a reticle to determine target identity and location. The instructor station uses a camera in a fixed location to provide an overall view of the range. Through a simple interface the instructor is able to control (start and stop) the targets and to determine the effect of a fire hit. The authors mentioned that by using this Augmented Reality system training becomes more effective and more realistic to live combat than by using traditional simulation systems.

The work of Franklin [Fra06] focuses on the application of Augmented Reality in the training domain. The test-bed is realized in the context of Forward Air Controller training. Using the system, the Forward Air Controller (trainee) can hear and visualize a synthetic aircraft and he can communicate with the simulated pilot via voice. Thus, the trainee can guide the pilot onto the correct target. The system can provide synthetic air asset stimulus and can support the generation of synthetic ground based entities. Positions and behavior of these entities can be adapted to the needs of the scenario. The author concluded, that the impact of Augmented Reality for training depends on the specific requirements of the end user and in particular on the realism of the stimulation required. According to the author, this is influenced by (1) the means of the required stimulation, (2) the criticality on how the synthetic stimulation is used, (3) the dynamism and complexity of the training environment and (4) the availability of a common synthetic environment.

An approach in the context of robot Programming-by-Demonstration is presented in [SZY*09]. In this system, the user is trained in acquiring the necessary skills for getting a deep understanding of the task to be programmed and a good comprehension of how to use the robotic system. The robotic system is based on a light-weight robot, which allows direct human-machine interaction and compliant motions. The system includes an Augmented Reality as well as Virtual Reality setup. In focus of training are two different sub-tasks of this system: avoiding robot singularities (i.e. "a condition caused by the collinear alignment of two or more robot axes resulting in unpredictable robot motion and velocities" [ANSI/RIA R15.06-1999]) and setting correct compliance parameters. Therefore the authors define learning accelerators, which are implemented and evaluated in both setups and present a training protocol for training Programming-by-Demonstration skills.

Also another work showed the potential of Augmented Reality in the context of Augmented Reality based human-robot interaction for industrial robots [BK04]. Interaction devices and tracking methods are analyzed in terms of cost, accuracy, etc. It has been pointed out that handhelds and monitor-based visualization devices are better suited for industrial environments than Head Mounted Displays, since they are more robust. Optical tracking methods were considered to provide best accuracy and flexibility, while promising rapid development at reasonable cost. The developed interaction prototypes concentrate on user training, programming and operation, and service and maintenance. The authors state that Augmented Reality has much potential for teaching trainees how to operate industrial robots.

2.2. Surgery and Rehabilitation

Also in the field of medicine the importance of Augmented Reality based training approaches is growing. Blum et al. [BHKN09] propose an AR simulator to improve the training of ultrasound skills. An ultrasound probe is tracked and the corresponding ultrasound images are generated from a CT (computed tomography) volume. Using in-situ visualization of the simulated ultrasound slice and the human anatomy, a deeper understanding of the slice and the spatial layout can be achieved. Both the trainees' and the experts' performance of an ultrasound task are recorded and synchronized in time. In an Augmented Reality based after action review the trainee can analyze the synchronized records. With this system training can be done without experts being present and feedback (replay of synchronized records) can be provided automatically. Thus, the two most important features of medical simulators, providing feedback and allowing repetitive practice [IMP*05], are covered by the system.

In the field of ultrasound simulators, there are also approaches that use a screen-based visualization. Weidenbach et al. [WWPR00] present an Augmented Reality based echocardiographic system in which a surface model of the human heart is linked with echocardiographic volume data sets. The ultrasound probe is tracked on a physical phantom using a magnetic tracking system and based on a virtual model of the heart the simulator generates the ultrasound image. The work of Maul et al. [MSB*04] addresses 3D ultrasound volumes from pregnant women, which have been acquired using a tracked ultrasound probe. The simulator was used for training of screening for normal and pathological findings, that is, to teach the measuring of the size of specific anatomy. Comparing the measurement errors to a control group that only received traditional training showed that the errors were much lower for the group that practiced using the training simulator. Hence, the authors conclude, that simulator-based training may provide an ideal tool to test, improve and monitor a physician's or technician's ultrasound skills in detecting fetal anomalies.

An Augmented Reality dental training simulator with a haptic feedback device is described in [RGH*10]. The simulator delivers kinematic feedback and hand guidance. It provides extremely realistic conditions for the trainees by combining 3D tooth and tool models upon the real-world view using a Head Mounted Display. Thus, the trainees can practice surgery in the correct postures. Important features such as applied forces and tool movement that characterize the quality of the procedure are monitored by the system. Outcome and process features of the trainee are compared to the best matching expert. First evaluation results show a promising potential of this simulator for supplemental training.

In [PKT*09] a prototype for a multimodal skills training system for Minimally Invasive Telerobotic Surgery (MITS) is presented. The system is based on MiroSurge system developed by DLR [HNJ*08], [HOK*08], [HKT*10]). The training system consists of a surgical workstation providing force and 3D visual feedback, which is identical with the real workstation and a Virtual Reality environment, in which the robotic systems and the training environment is simulated. The VR environment can be adapted to the trainees capabilities and the training situation (e.g. the difficulty level can be configured). The focus of the skills trainer is to transfer the skill required for operating the involved telerobotic system to surgeons. Skill acquisition can be performed in a Virtual Reality scenario without instead of using the complete MITS system on animal cadavers. The selected training task is identical to state-of-the-art surgical training for laparoscopic surgery. Besides the description of the system, a training protocol and evaluation scheme is provided in this work.

A training system for the transfer of skills relevant for in maxillo facial surgery tasks is shown in [GMBT09]. The training system consists of a multimodal training platform featuring involving visual and audio feedback. The authors intend to provide a pedagogical training system allowing to ease the transfer of skills from senior operators to novices for delicate operations. The key components of the training platform are the haptic devices, which allow to capture the hand movements of the surgeon and to render the tool-bone interactions. A vibrating handle was developed in order to allow the reproduction of the high frequency vibrations produced by the surgery drill. For the second hand holding the retractor or vacuum cleaner a second haptic device, mainly used for mouth tissues interaction, has been integrated. Visual feedback is provided on a screen positioned above the simulated operating site. Also surgery sound is synthesized and replayed. The authors also specify preliminary learning accelerators and exercises to be evaluated in further work.

In [LFA*09] a series of haptic-based virtual applications for the rehabilitation of upper-limb stroke patients. The developed rehabilitation exercise tasks refer to a novel multimodal interface (Bimanual Haptic Desktop System). This interface integrates the haptic and video display functionalities within the work plate of a desk. The aim of the proposed video-game-like exercises is the recovery of basic moving skills and moving steadiness of the patient's upper-limb. Furthermore, the exercises provide continuous measurement and evaluation of performance data to analyze the interaction between the subject and the environment. The authors highlight three main benefits of the system, namely (1) a coherent and colocated rehabilitation involving force stimuli and visual information; (2) a greater choice of rehabilitation sessions through the combination of three different exercises; (3) a convenient graphical user interface.

2.3. Sport and Entertainment

An example for a multimodal training platform in the field of sport is given in [SGTRF09]. The paper presents a study with focus on skills transfer of boxing. The developed training system includes Virtual Reality environments, tracking systems, vibration stimuli and human force measurements (EMG). The aim is to analyze and train novice boxing trainees in order to support them in the acquisition of an appropriate boxing technique. The multimodal platform is based on two training methodologies, force and technique. Based on a neuromechanical analysis of movements and punches, visual-tactile feedback strategies have been specified and applied in order to enhance the trainees perception. Thus, the learning process of the trainee is accelerated.

A training system for skills transfer in rowing is presented in [RFF*09]. The system consists of a mechanical platform, a set of sensors, a system for data reading, elaborating, and storage, and a screen. The mechanical platform allows the trainee to experience motion and force feedback similar to conditions of a real boat. The visualization component provides two types of visual feedback, realistic visual rendering of the rowing environment and specific training elements like guiding lines and performance indicators. The simulation model of the dynamics is composed of three dynamic models: one for the boat, one for the human, and a last one for the oar. The model can be adapted by changing different boat and oar parameters to provide different training conditions.

In [LZA*09] an approach for training of juggling skills is described. The developed training platform focuses on timing and spatial constraints (kinematics) of juggling. The position of the jugglers hands is captured using a Polhemus tracker. The juggler is holding a sensor in each hand by which he can control the ball. An analysis system estimates the initial skill level of a juggler by comparing captured data of the trainee with a database of jugglers with various skill levels and choosing the closest match. During the training, the speed of the ball and the trainees' emphasis level (effort, attention allocation) vary. The authors propose learning accelerators and training protocols to be further evaluated in subsequent work.

A very early example of an Augmented Reality training system addresses the training of billiard tasks [JEW*97]. The system computes the best shot for a given image of a pool table, by locating the pockets and the balls and computing the easiest shot that can be performed. The computed motion paths (for the ball to shoot and all other relevant balls) are projected in the view of the Head Mounted Display the user is wearing. and overlaid onto the table. The detection of pockets and balls is realized using computer vision techniques.

An Augmented Reality training system for batsman is shown in [Fli10]. The aim of the training platform is to improve timing, reflexes and batting skills for any level of batsman. The prototype system addresses in particular cricket, but it can also be adapted to any bat and ball sport like baseball or softball. The system architecture consists of two vision sensors, an accelerometer and a graphics engine. Feedback and interaction are provided in real-time. The motion of the bat is captured using 3D object tracking.

Also in the field of entertainment the use of Augmented Reality based training is in focus of some research groups. In [CBC03] an Augmented Reality based learning assistant for playing electric bass guitar is described. There two key elements of the system: tracking of the users' fingers and the guitar, and overlaying graphical annotations synchronized with sound. The position of the users' fingers and the guitar are determined using a marker-based tracking (ARToolKit [KB99]). Thus, notes written on a sheet of music can be marked on the guitar. The main goal of the system is to accelerate the process of associating a musical score with the sounds represented by that score and with the fingerboard. The authors asses, that perceptual and cognitive discontinuities, that occur during conventional learning of instruments, can be eliminated by dynamically overlaying appropriate information onto the instrument.

2.4. Summary

It becomes obvious that existing skill training systems focus either on technological aspects such as the capturing of forces and movements and the development of simulation models and rendering techniques, or on the training of the underlying motor skills. Most of the approaches provide a very strong guidance for the trainee and thus embarrass the trainee's active involvement in the acquisition or enhancement of skills. In this way, the user gains the motor ability for performing new tasks, but he does not acquire the underlying skills that are necessary to transfer the learned training tasks to similar situations and to efficiently acquire new tasks. To train the skills that underly certain tasks (e.g. maintenance tasks, surgery tasks) the cognitive aspects of skills must be taken into account and addressed in the training system. In this thesis an approach for identifying the relevant cognitive components and integrating them into the training process is presented in order to improve traditional training systems.

3. Skill Acquisition and Training

This chapter addresses basic theories and methodologies of the acquisition and training of skills. Aspects that are relevant for the development and the evaluation of skill training programs have been compiled and will be described in the following.

3.1. Definition of Skill

Learning from experience and adapting the behavior is the most remarkable ability of human organisms compared to all remaining organisms and distinguishes them clearly. How individuals differentiate in the performance of any given task depends mainly on practice and experience. Human skills comprise manifold categories such as perceptual (visual, auditory, touch), memory, manipulation of the abstract (language, mathematics, music, logic, etc.), motor, procedural and more. Skills differ from talent, since talent seems native while skill is learned by doing and interacting with the environment [PRSGA*10].

The word *skill* has different meanings: the origin is based on the late Old English word *scele*, which means knowledge, and from the Old Norse word *skil*, meaning discernment and knowledge ([SKI07], [PRSGA*10]). To give a general definition, skill can be understood as "the learned ability to do a process well" or "the acquired ability to successfully perform a specific task" [SKI07]. The term *task* means the elementary unit of goal directed behavior ([Gop04]) and is also a fundamental concept in the study of human behavior, that is strictly connected to skill. Moreover, skill is not associated only to knowledge, but also to technology, since technology is the study of skill (literally in the Greek) [PRSGA*10].

In the context of this work skill can be simply defined as an activity of a person involving a single or a group of movements performed with a high degree of precision and accuracy [Sin80]. This means, *skill denotes the capacity acquired by learning to reach a specified goal in a specific task with the maximum of success and a minimum of time, energy, or both* [BDL*10].

Based on this simple definition it becomes apparent that skill cannot be considered as a general and abstract ability, but rather as a specific and learned competence acting in an limited ensemble of situations [SKI07]. It can be acquired by practice and movement-related information and can be initiated and refined by demonstration. For evaluating the level of skill a number of criteria which have been defined and analyzed by Bardy et al. [BDL*10]. Some of them are listed below:

Accuracy of the outcome: The accuracy of outcome with respect to the assigned goal is the most often used skill measure. It is measured by analyzing errors to the targeted skill performance (spatial and /or temporal).

Consistency of responses: The outputs of skilled movements are characterized by stability and reproducibility. A traditional method for measuring stability is assessing for example the standard deviation of a set of successive outcomes. It has also been suggested to measure not only the amplitude of variability, but also its temporal structure. That is, consistency does not necessarily denote a "behavioral stereotype", but rather a behavior characterized by "repetition without repetition" (i.e. the experts' ability to regulate small details in the performance of actions in order to achieve a stable performance [Ber67]).

Efficiency: This denotes the ability to reach the desired goal at minimal cost [BDL*10]. It refers to diverse levels, such as the cognitive level (the use of automatic processes allowing a decrease in mental load) or the metabolic level (a reduced metabolic cost in expert to reach the same performance) the neuromuscular level (reduction of co-contraction and more phasic muscular activation in experts), which are not independent from each other. As skill acquisition progresses these various levels generally show the same tendency toward economy.

Flexibility and adaptability: Skilled behavior distinguishes itself by the fact that it can cope with endogenous and exogenous uncertainties. This indicates that skill is not specific to a dedicated task, but rather to a set of similar tasks. Thereby the fundamental problem of skill generalization and transfer is raised. Flexibility and adaptability are crucial and complementary aspects to the stability properties described above. Even if variability is often functional, as a consequence of continuous and functional adaptations it is also operating at the level of perception-action loops [BDL*10].

Coordinative structure: A skilled movement is never performed in the exact same way, and always possesses a certain degree of variability. This is because skills are characterized by a delicate spatio-temporal organization of sub-movements. Those ensembles include task specific synergies among body segments, joint angles and end-effectors and also synergies between the agent and the environment. However, there is a pattern of order which is common for each sample. The coordinative structure must be able to cope with perturbations (internal and external) in order to achieve a good level of proficiency. That is, skilled behavior is characterized by stability. Hence, the elaboration of a coordinative structure is another important factor for evaluating skill.

3.1.1. Skill Decomposition

As already mentioned, the acquisition of skill is accompanied by various transformations in coordination and perception-action, which involve sensorimotor and cognitive skill elements [BDL*10]. Thus, each specific skill (e.g. manipulating, assembling, juggling, rock climbing, etc.) can be decomposed into a set of sensorimotor and cognitive components, which are also called *sub-skills*. Generally speaking, *sensorimotor* skills refer to the correlation between and motor components and perceptual components, while *cognitive* skills concern higher-level cognitive activities that orient, formulate, monitor and regulate the sensorimotor performance [BDL*10].

Sensorimotor Sub-Skills	
Balance and postural control	The regulation of balance (e.g. static, dynamic) and posture (e.g. muscles, joints) allowing for a successful achievement of the manual performance. Performance evaluation: inter-segmental (e.g. elbow, hip) and inter-muscular coordination (e.g. flexion, extension), center-of pressure variables.
Bimanual coordination	Bimanual coordination is the functional synchronization (space and time) of arms, hands, and fingers. Performance evaluation: relative phasing between the coordinated elements and its stability.
Hand-eye coordination	The synchronization of eye, gaze, and effector relating to the main information that is perceptually detected. Performance evaluation: gain, relative phase, coupling variables (between eye/gaze/hand).
Interpersonal coordination	The coordination between two or more people evolving from a multilevel combination of components such as motor dynamical principles, neuroscience constraints, and sociality. Performance Evaluation: relative phasing between persons and its stability.
Perception-by-touch	This refers to the perception via the haptic modality. It is based on various receptors embedded in the skin providing information about temperature, pain, and mechanical properties (roughness/vibration/compliance). Performance evaluation: evaluation by psychophysical methods, e.g. absolute threshold (smallest identified intensity), just noticeable difference (smallest identified difference in intensity), sensitivity (level of ability to distinguish between different intensities), etc.
Prospective control	The ability to anticipate the prospective place-of-contact and time-to-contact based on spatio-temporal optic, acoustic, and haptic information. Performance evaluation: movement initiation and movement-information coupling variables (e.g time-to-contact).
Proximo-distal coupling	The spatio-temporal coordination of proximal, gross components with distal, fine, manipulatory components (e.g organization of the body underlying arm movements, synergy between arm postures and hand movements). Performance evaluation: relative phasing between proximal and distal movements, cross-coherence, cross-correlation, etc.
Respiratory-movement coupling	The coupling/synchronization of breathing and movement allowing for efficient performance. Performance evaluation: phase and frequency synchronization patterns and their stability, amplitude.
Fine motor control	The fine regulation of force and movement (e.g. internal forces applied on the surface) allowing for a successful performance (e.g. of drilling, pasting) with the required level of accuracy. Performance evaluation: force/positions accuracy, trial-by-trial variability, etc.

Table 3.1.: Classification of sensorimotor sub-skills (adapted from [BDL*10]).

Cognitive Sub-Skills	
Control flexibility & attention management	The ability to change performance strategies and response modes, to apply and manage new attention policies, and to pursue new intentions and objectives. Performance evaluation: adaption to varying task demands and to changes in attention allocation.
Coping strategies & response schemata	This is defined as a vector of importance or attention weights. The weights are calculated over the sub-elements of a task, which refer to a dedicated (performance) objective. Performance evaluation: collection (number and type) of strategies to cope with changes in task demands and intention.
Memory organization, structure & development of knowledge	Expertise to well formulated and organized multi hierarchy, and of task specific memory and knowledge bases facilitating to encode, to retrieve, and to execute performances. Performance evaluation: speed/accuracy of encoding, response, and decision making; number/multiplicity/speed of generating alternative solutions for the task performance.
Perceptual-observational skills	The ability to detect, test, extract task relevant information in/from the environment, and to identify static patterns and dynamic regularities. Performance evaluation: speed/efficiency of detecting relevant information, speed/accuracy of response, recognition and use of redundancies, etc.
Procedural skills	The ability to follow repeatedly a sequence of actions step-by-step in order to reach a specified goal. Procedures are sequences of ordered activities that need to be executed in the performance of tasks. The performance of each task can be subdivided into a number of procedures/activities, which have to be repeated each time the task is performed (e.g. procedures of preparing coffee). Performance evaluation: speed/efficiency/completeness of performance, transition smoothness between steps within a procedure, number of performance errors within a procedure.

Table 3.2.: Classification of cognitive sub-skills (adapted from [BDL*10]).

Within a cooperation with human factors and cognitive scientists several sensorimotor and cognitive skills have been identified to cover training approaches in the fields of *sport & entertainment* (Chapter 2.1), *surgery & rehabilitation* (Chapter 2.2) and *industry & manufacturing* (Chapter 2.3). Among a virtually infinite number of sub-skills that compose human activities, they have been selected because of (i) their key importance for the successful achievement of the skilled performance, (ii) their coverage of complementary perceptual modalities or effectors, (iii) their visible evolution over time and learning, and (iv) their anchorage in a solid state-of-the-art in basic human movement and cognitive sciences. These sub-skills are specified in Table 3.1 and Table 3.2. The tables also provide information about relevant elements for evaluating the performance of the sub-skills.

As a result of the natural embodiment of cognitive phenomena into sensorimotor dynamics, what is an important factor in this work, the two categories (sensorimotor and cognitive) are largely interdependent. Thus, the differentiation between cognitive and sensorimotor skills is partly arbitrary. In general, the term embodied cognition is largely understood in the action-based context, while off-line cognition, what means cognition decoupled from the environment, is rather body-based. Accordingly, if no task demand is present, hidden motor representations can be evoked by perceptual inputs (such as vision or touch). For example, judgments of whether a screwdriver is screwing or unscrewing are faster when the orientation of the handle is consistent with the manual dominance of the observer [dS97]. Also brain imaging studies have shown, that the perception of visually presented objects affords actions that can be made towards them, irrespective of the task itself [GD02]. Another interesting fact is that by observing an action individuals automatically create an internal replica of the corresponding actions in their premotor cortex [BBF*01]. From this it can be deduced that embodied cognition plays a role in representing and understanding the behavior of other individuals [Wil01], which is used, for instance, in learning through imitation.

Together, these insights suggest that perception is not mere a perceptual process which precedes symbolic representations of actions to be performed, but rather is embedded with cognition. Perception and cognition link body and goals to the environment ([VTR91], [Pro06]). What and how one perceives in the world is affected not only by inputs like visual information, but also by his intentions, emotions, and physiological state. Thus, the classification presented in Tables 3.1 and 3.2 is only a convenient way to facilitate elaborating technological tools for improving the learning of these sub-skills.

3.1.2. Primary and Secondary Skills

In general, human skills can be classified into primary or secondary skills. *Primary* (elementary) skills determine the flexibility of a humans body, the sensory systems of a human and his mental capabilities [SKI07]. They can be for instance perceptual, motor, procedural and so on. Examples are walking, hand and finger coordination, speaking and identification of tones. Using perceptual skills sensory data can be interpreted and organized into objects, categories, and events (e.g. in words). Motor skills induce the activation of the muscles for controlled and coordinated movements. Procedural skills are characterized by implicit knowledge of how to do things. Those primary skills are the building blocks of secondary skills.

Secondary skills build on primary skills. They combine and integrate several elementary skills to develop the ability of the human of performing a dedicated task [SKI07]. Examples are typing, playing music, car driving, or playing football. The acquisition of those secondary skills requires the existence of many primary skills of the trainee. The relationship between primary skills and secondary skills is a many-to-many relation. Each secondary skill comprises many primary skills and each primary skill can be a part of many secondary skills. For example, driving a car (secondary skill) requires the ability of reading signs, geographic knowledge, knowledge about rules of the road, many perceptual and motor skills, and more. Accordingly, secondary skills are a composition of sensorimotor components and cognitive components and of the mutual influence of both components on each other [SKI07].

3.2. Models and Theories for Skill Acquisition

For the development of efficient skill training programs it is important to understand and analyze the human's skill acquisition process. Research on cognitive and motor skills has pointed out that the process of skill acquisition differs from the execution process of the skill. The main difference is that learning is a gradual process that happens over many performance attempts, resulting in behavior that suffers less from transitory factors such as fatigue and anxiety [GB03]. Studies about learning and performance of specific tasks resulted in knowledge on practical issues, such as the optimal scheduling of practice (e.g. amount, duration, and frequency). Relevant theories and models describing this learning process will be discussed in the following.

3.2.1. The Three-Stage Model Theory

Fitts [Fit54] developed a three-stage model of skill acquisition, which has gained a great amount of support in further research and practical literature (e.g. [NR93], [SM04], [AZ97]). In this model learning is described as a continuous process in which the model of processing information changes gradually during the learning progress. According to Fitts, the development of skills proceeds through three stages: a cognitive stage, an associative stage and an autonomous stage (see Figure 3.1).

During the initial *cognitive stage* the learner acquires basic understanding of the requisite behavior (e.g. the movement) through instruction and observation. Using this information he is able to produce a rough approximation of the skill. At this stage, the learner experiments with different configurations of the execution (e.g. movement changes) and hence the performance is variable and may be defective. In the next stage, the *associative stage*, the behavior patterns (e.g. movement patterns) are refined and motor programs for task execution begin to develop. Initial errors in understanding the task are eliminated and appropriate methods for proceeding are detected. The lengths and amount of practice in this stage depends on the learner's abilities and the complexity of the tasks. The last stage is the *autonomous stage*. To achieve this stage, extensive practice is required, which can extend over a period of many years. The performance of the skill becomes increasingly integrated, rapid, and autonomic, and accordingly the performance is incrementally refined. In this stage, the learner can perform the skill with minimum effort and few errors.

Anderson developed a similar three-stage model [And82]. In his theory, the model distinguishes between two types of knowledge relevant to skilled performance: declarative knowledge (*what* knowledge), and procedural knowledge (*how* knowledge). The declarative knowledge comprises factual knowledge about a skill. The procedural knowledge describes knowledge about how to carry out various tasks and activities. According to Anderson's theory, the basic process of skill acquisition starts out in *declarative stage*, in which facts about the skill domain are interpreted and applied. In the second stage, called *knowledge compilation*, this declarative knowledge becomes compiled into a procedural form and is directly embodied in procedures for performing a skill. That is, it is directly applied without the intercession of other interpretive procedures. The process of knowledge compilation consists of

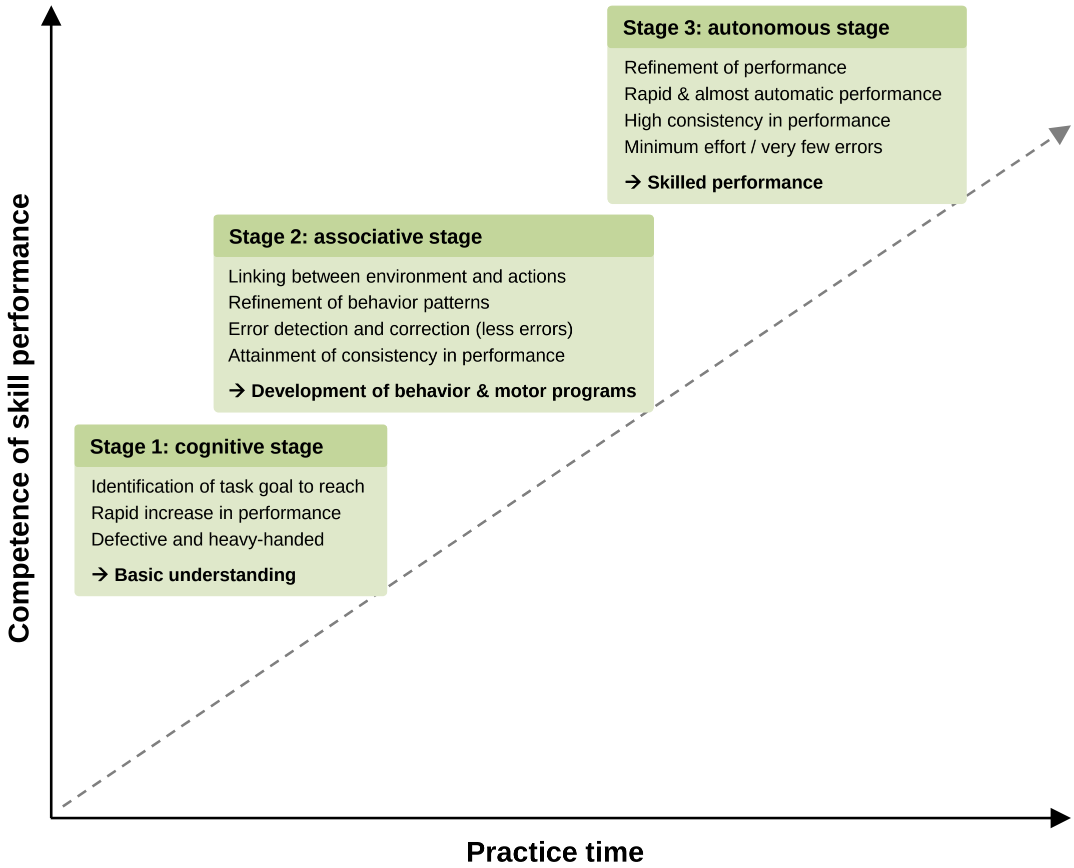


Figure 3.1.: Fitts' three-stage model of skill acquisition: a continuous process with gradual changes in the nature of information processing (taken and adapted from [DBB08]).

subprocesses of composition, which collapse the sequence of productions into single productions, and proceduralization, which embeds factual knowledge into productions. In a third stage, the *procedural stage*, this procedural form undergoes a continual process of refinement of conditions and gradual increase of speed. This stage analysis provides an approximation to characterize a complex system of interactions.

3.2.2. The Dreyfus Model

Also the model developed by Dreyfus [DD80] indicates that when individuals acquire a skill through external instruction, they normally pass through several stages. The Dreyfus model states, that in the acquisition and development of a skill, a person passes through five levels of proficiency: novice, advanced beginner, competent, proficient, and expert. These five levels reflect changes in three general aspects of skilled performance. The first one is a shift from reliance on principles to the use of past concrete experience as paradigms. The second is a change in the learner's perception of the demand situation, in which the situation is less a compilation of equally relevant bits, but rather a complete ensemble in which only certain parts are relevant. The third level is a transit from detached observer to involved performer. The performer no longer stands outside the situation, but is now engaged in the situation ([Ang02], [Ben04]).

In the stage of the model, the *novice* stage, novices are taught rules in order to help them performing. The rules are context-free and independent of specific cases, so they tend to be applied universally. The novices follow the given rules with no sense of responsibility beyond following the rules exactly. This rule-ordered behavior typical of novices is extremely limited and inflexible.

Advanced beginners can demonstrate marginally acceptable performance. They have coped with enough real situations to note the recurring meaningful situational components. These components require prior experience in actual situations for recognition. The learner begins to formulate principles to guide actions, which are based on experience.

Competence develops when the learner begins to see his actions in terms of long-range goals or plans of which he is consciously aware. That is, the learner develops organizing principles to quickly access the particular rules that are relevant to the specific task. A *competent* learner is characterized by active decision making in choosing a course of action. The competent learner does not yet have enough experience to recognize a situation in terms of a general picture and can still not decide which aspects are most important.

Characteristic for a *proficient* learner is, that he perceives situations as wholes rather than as separated parts or aspects. He develops intuition to guide his decisions and devise his own rules to formulate plans. Once the proficient learner has a deep understanding of the situation overall, however, the formulated rules provide direction as to what must be taken into account.

Finally, the *expert* no longer relies on an analytic principle (rule, guideline) to relate his understanding of the situation to an appropriate action. The expert operates from a deep understanding of the total situation. The expert is no longer aware of features and rules. His performance becomes fluid, flexible and highly proficient, what does not mean that the expert never uses analytic tools. Situations with which the expert does not have any prior experience require highly skilled analytic ability. Analytic tools are also necessary, when the expert gets a wrong comprehension of the situation and thereupon notices that resulting events and behaviors are not as expected. In case that alternative perspectives are not available to the performer, the only way out of a wrong comprehension of the problem is trying to solve the problem analytically.

3.2.3. Information Processing Theories

A popular theory postulates that perceptual-motor information can be represented in the central nervous system (CNS) [Woo99]. Such representations are developed through learning. As a result, in the CNS a set of motor commands are stored that control movement behavior. A basic assumption is that the brain acts like a computer to process information and to produce outputs in behavior. Information processing is based on a series of discrete cognitive stages including perception, decision making, and response execution [Kee68] (see Figure 3.2). Usually, reaction time is used as key variable for explaining the stages of information processing in tasks that vary in complexity.

It has been pointed out that feedback plays a major role in information processing. The *closed-loop theory* [Ada71] implies that normal movement is regulated by a continuous comparison between current sensory information and information resulting from successful movements (see Figure 3.3). That is, feedback is used to control states or outputs of a dynamical system. By using result-related feedback (knowledge of results) learners in early learning phases develop representations of the expected movements (perceptual traces) to estimate their progress. In the closed-loop theory there is no direct connection between the output of the system and the actual conditions encountered. Hence, the system does not and can not compensate for unexpected sensory information [KG02]. It became apparent that mainly slow positioning movements (e.g. cycling, steering a car) fit better the close-loop theory than the open-loop theory, in which no sensory feedback is available (see Figure 3.3). A common key element for both theories is the capacity of the CNS to store a huge number of representations to control many different movements needed in daily life.

The *schema theory* [Sch75] combines aspects of both open-loop and closed-loop control. In this theory, a schema is set of rules that concern the execution of a movement response, which is related to feedback received during and after performance. The goal is that the learner acquires a robust schema that can produce functional movements for a particular class of actions (e.g. driving a car) under varying performance conditions. Therefore, variable practice conditions are necessary.

3.3. Training Methodologies

When designing a training program it is useful to consider training methodologies that have proven their worth. The methodologies which are most relevant for this work are summarized below.

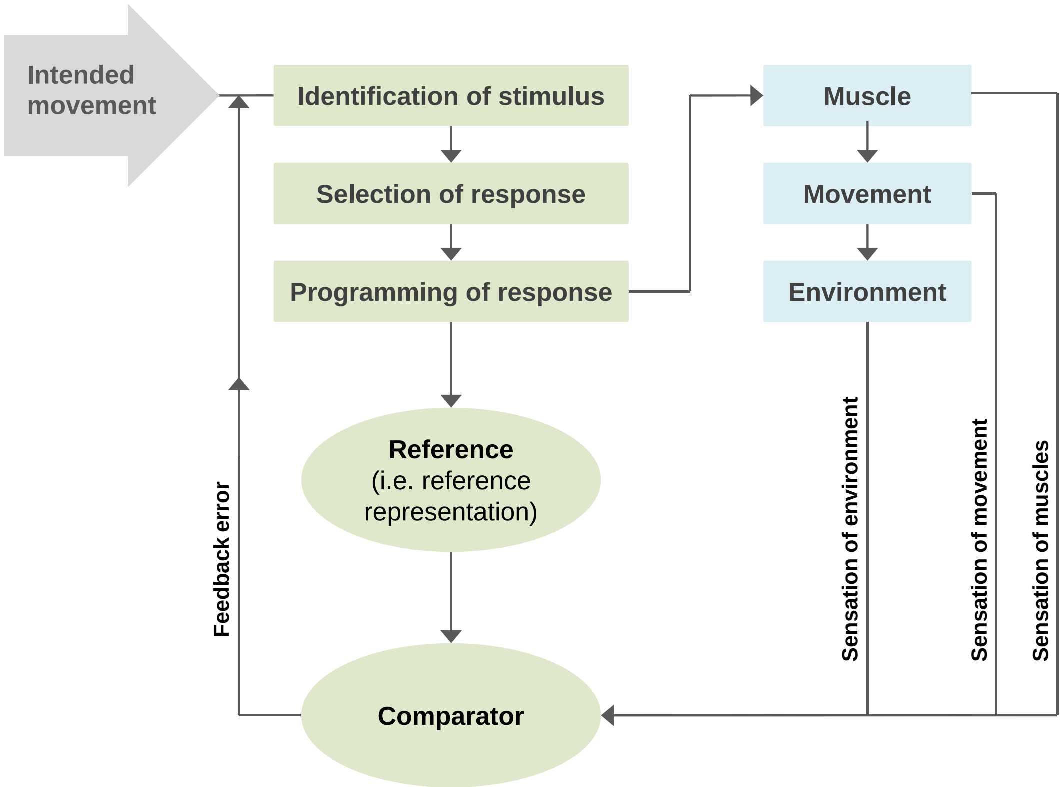


Figure 3.2.: Information processing in the CNS: When information is processed a series of cognitive stages is involved, such as perception, decision making, and response execution (adapted from [DBB08]).

The main objective of a training program is to maximize the trainee's task performance for a given task. The focus can be either to bring the trainee to develop the ability to perform a given task accordant to pre-specified performance criteria and goals (e.g. performing successfully a maintenance procedure of the engine of a car), or, if task performance criteria can not be specified or if they are set as highest possible levels, the aim is to optimize the trainee's performance of a task (e.g. become a competitive rower). While in the first case the focus is on achieving the specific goals of the task, the aim of the second case is to maximize the trainee's capabilities.

A lot of work has been done in theory and research on training, development of training protocols and design of training platforms. In the following the state-of-the-art of the training aspects, which are most relevant for this work, will be summarized. These aspects comprise the following topics:

Consistency the fundamental and most critical requirement of learning and skill acquisition

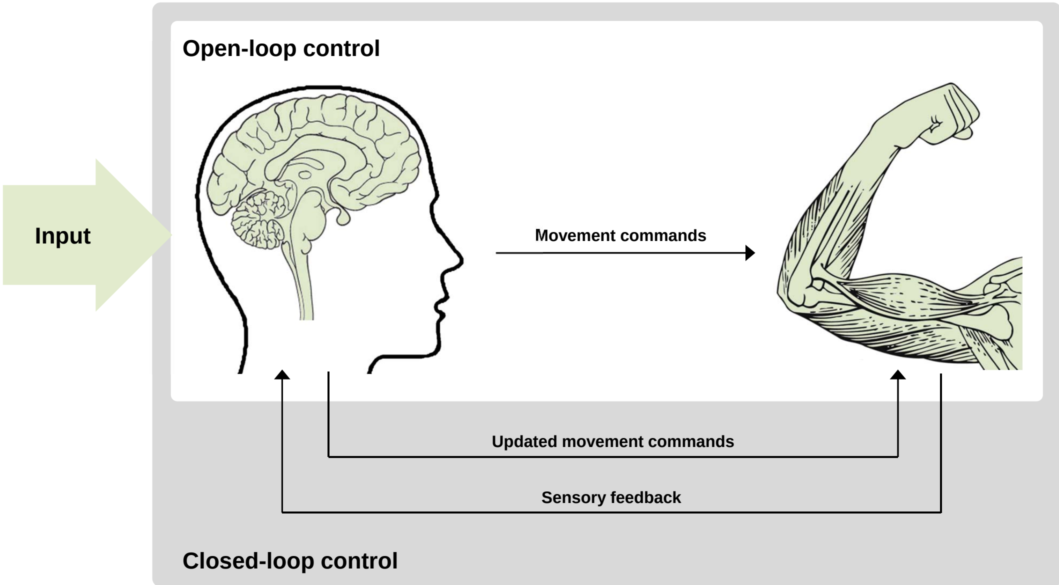


Figure 3.3.: Open-loop and closed-loop control: a closed-loop control system comes with an additional feedback channel.

Modes of Experience spontaneous practice (trial and error) compared to guided instructions and coaching

Feedback the role of feedback and knowledge of results

Discovery and Observational Learning how active exploration and observation influences learning

Variability the use of variability during training

Partial Training Methodology the effect of partial task training for complex skills

3.3.1. Consistency

In the acquisition process of experience and expertise the human's interaction with the task and the environment when the situation reoccurs is a crucial factor. Successful responses should be repeated and failed responses should be discontinued. From this it follows that consistency in the mapping between actions and results is an important factor for learning and training [SKI07]. Humans learn about their environment from experience and they learn how to act from successfully performed actions in their past behavior. Through a consistent mapping between actions and results they can learn which actions are suitable in order to reach a specific goal. Such a cause-effect relationship cannot be evolved without consistency [SKI07]. In this case, the human may learn that there is nothing he can do in order to reach the designated goal and therefor he does nothing, what is called "Learned Helplessness" [PMS95]. Through Learned Helplessness the human is put a state of freeze and despair what involves further disturbances [PMS95]. For instance, if parents at one time permit a special behavior and forbid it at another time, children may be at a loss.

Schneider and Shiffrin showed that without a consistent mapping between elementary stimuli and response units of a task the progress of skill acquisition is very poor or not obtained at all [SS77]. Their experiments revealed that a consistent mapping between stimuli and responses (CM) results in automatic, efficient behavior of the trainees, while inconsistent variable mapping (VM), in which the mapping rules were changed, yields a very poor improvement of the behavior. Two experimental search tasks were conducted in which subjects were shown letters on a computer screen in a high display frequency. The setup of the experiment was as follows: Each display screen ("a frame") included four characters arranged to form a square around a central marker ("a stimulus"). The subject had to judge whether a target letter belonging to a pre-specified group of letters had been presented. During the experiment the frame display duration (time from starting time of displaying one frame to the starting time of the next frame), frame display size (number of letters in the frame), memory set size (number of possible targets) and consistency of target- distractor mappings were manipulated. In case of consistent mapping (CM), the targets and distractors were distinguished by category (e.g., letters or numbers), hence, target was always target and distractor was always distractor. In the variable mapping (VM) trials, targets and distractors were from the same category, so target was sometimes a previous distractor and present distractors could be previous targets.

The main results of the experiment demonstrate that the subject's performance in the VM trials was highly dependent on memory set size and frame display size (number of targets and number of letters in the frame), while performance in the CM trials was largely independent of these variables. Overall the results show that performance in CM condition was much better than in VM condition. The end point for the CM function (120 ms), when performance was best, was only the starting point for the VM function (i.e. the minimal performance levels) [SKI07].

It can be concluded that the human organism is very sensitive to any type of regularity, since he uses those regularities to improve the learning performance. Hence, consistency is a crucial factor for developing learning and training protocols.

3.3.2. Modes of Experience

Different studies showed that unguided experience and spontaneous practice by trial-and-error may not maximize the trainee's performance and bring him to highest levels expertise ([DFS89], [Gop93]). Although the accumulation of experience in task performance forms the basis for learning and acquisition of a task, limited knowledge on the nature and best heuristics of the performed task and limited exploration and exposure to task variants [HP91] may lead to inferior learning in spontaneous practice as compared to guided experience. The same applies for deficient monitoring of self performance and missing feedback and knowledge of results. Compared to unguided experience, the existence of instruction is even more important than the selected method of training itself.

Various studies which compare learning strategies for complex high demand tasks [DFS89] have been conducted. One example is the work of Gopher et al. [GWS89] who carried out a study at the example of complex computer game. The game comprised difficult dynamic and discrete manual control, visual scanning and search, short-term and long-term memory demands, all performed under severe time constraints and attention load [FBG*89]. The subjects, who had to practice the game in ten one-hour practice sessions, were divided into four groups. One group (the control group) played the game during the ten sessions without any guided training, just general instructions were given. The three remaining groups received guided training during sessions 2-6 and played the game without any guidance on session 1 and on sessions 7-10. The results showed indeed that after the ten practice sessions all groups improved with training, but that the performance of subjects in the three guided groups (and most pronounced in the double manipulation group) was significantly better than the performance of the control group participants. Moreover, the results provided evidence indicating that the ultimate differences will be much larger between guided and unguided training [SKI07].

Another study, which explored a more theoretical and methodological approach to training based on research about Hierarchical Task Decomposition showed similar results for instructed and uninstructed training groups [FW89].

To conclude, the results of these and other similar studies validate the significance of guidance, instructions and coaching. It can be stated that guided practice has an enormous impact for the acquisition of complex tasks.

3.3.3. Feedback

As mentioned in Section 3.1 the ability to adapt the behavior based on previous experience in order to achieve a better task performance in future is the most remarkable ability of human organisms. From this it follows, that helping the organism to adapt to his environment is a crucial factor in learning and skill acquisition. In order learning can take place, feedback and knowledge on the consequences of performance are mandatory requirements. In general, there are two main types of feedback. Internal feedback and knowledge of results [SKI07]. Internal feedback is derived from the performed action itself (e.g. kinesthetic and proprioceptive feedback in the performance of movement). Knowledge of results (KR) informs the performer on the match between his action and the designated goal. For example, when throwing a dart, the performer receives feedback from his muscles and joints about the force, direction and speed of the throw, the hit point of the dart on the board and more. This information provides KR about the quality of the throw relative to the board center. If a clear objective target does not exist (e.g. task is to lift hands or to run faster), it has to be defined by an external source, such as an instructor or a coach. This means, that KR has to be provided by instructors. In this case KR implies matching of criteria between trainer and trainee [SKI07].

KR can be given in a very different mode or modality than the performed activity. KR can be concrete (e.g. verbal, aural, written, visually displayed) or abstract (+,-) [SKI07]. Even while a task itself is for example sensorimotor or visual, nevertheless providing KR is highly effective. This suggests the human's capability of using the abstract or semantic information given in one modality to adjust representations and consequences of behavior in another modality.

The role of feedback and KR for learning is well-known for many years and has been explored extensively in various studies ([Ski06], [TC32], [Mag07], [SYSS89]). The work of Skinner et al. [Ski06] for example discusses the role of feedback and reinforcements in learning of animals and humans and accounted for many attempts of training and behavior shaping techniques. Trowbridge and Cason conducted another early study which demonstrated the direct relation between the subject performance of simple motor movements and the rate and quality of the KR that they received [TC32]. In [Mag07] and [SYSS89] a good overview of the evolution of research in this problem area over the years is provided.

3.3.4. Discovery and Observational Learning

One would expect that guiding a trainee passively using direct manipulation techniques with artificial devices (e.g. force feedback devices, exoskeletons, etc.) or with strong passive stimulation of the targeted movement might be a promising approach for skill training. But passive guidance and strong passive stimulation have rather proven to be inefficient [New91], since learning requires an active involvement of the trainee. He needs to discover the regularities and changes of the task by himself. Indeed, manipulating the amount of information to be processed and "channeling" the behavior of the trainee to the optimal task execution can enhance learning. Such a channeling can be realized for example by providing information about the task management. But anyway the trainee should always be able to actively explore the task and the environment ([New91], [SKI07]).

To determine the way how to guide the trainee toward a successful performance, a task analysis is required. This task analysis can be done in the cognitive or in the perceptuomotor context or both. The formation of movement solutions is based on various constraints, such as the identification of key variables, degrees of freedom at work and more [HDBB97]. Therefore, the sustainment of the natural task properties and the facilitation to discover them should be considered in the design of training interfaces.

When the trainee has the possibility to search for and to find out the system dynamics that organize information and action in specific tasks, training is often efficient. This experience of discovering different solutions for the task (whether successful or not) allows the trainee to explore and exploit the dynamical laws, what is a crucial factor for learning. Thus, technological interfaces for training should ideally encourage, constrain, and circumscribe variability, but not totally prevent it. Furthermore, there is usually not a unique route toward expertise. Different individuals often take different routes.

In the approach of learning by observation the performer has the opportunity to watch the responses of another performer. An example for observational learning is person, that acquires knowledge about how to play football by watching a football game on TV. According to the fact that many components of a complex skill may be of cognitive nature (e.g. game rules, tactics and strategies) the question raises, if cognitive components can be acquired without active engagement in task performance. Various studies pointed out passive observation can produce good cognitive representations of skilled actions

([Ban86], [CB90], [SGR98]). But nonetheless, being active in learning is considerably better and different from being passive.

3.3.5. Variability

An inherent property of complex tasks is that they are variable in difficulty, workload, and demand [Gop07]. This variability can arise from differences between several trials (i.e. from trial to trial), several events (i.e. from event to event), or scenarios. During training the trainee should learn to adjust his response to cope with such changes. The introduction of variability during training can enhance learning and improve the trainee's generalization of the task to changed conditions. Two types of variability can be distinguished, bottom-up variability and top-down variability [SKI07].

In *bottom-up variability* different single stimuli or single responses vary in their difficulty or demands. An example for bottom-up variability tasks is the addition of two digits, where the difficulty of the single task depends on the digits to add. In case of *top-down variability* the conditions of the task vary, its basic properties are not changed. Top-down variability is created by the performer, who determines the conditions for the task. For example, a performer may adopt different attention policies and apply strategic considerations etc. [Gop93].

Several studies showed that the introduction of top-down variability can be a powerful training technique in the training of complex skills ([Gop93], [Gop07]). Furthermore, it was pointed out that applying top-down variability in training can improve the performance, the adaptation to changes, and the generalization to new conditions.

One approach of introducing variability in training is changing emphasis. Under the training protocol of emphasis change subjects are instructed at different task sessions, to change their priorities for the performance of major task elements. This protocol turned out to be highly effective in improving subject competence in the performance of complex task ([GWB94], [Gop07]).

Another approach for integrating variability into training is the introduction of a secondary task, which can be used as a tool to force subjects to explore alternatives for the task performance by making them change their response and coping strategies with primary task demands ([HP91], [YEYG04]).

3.3.6. Partial Training Methodology

As already mentioned, a complex skill consists of many task segments that must be performed. In traditional training protocols the task is decomposed to segments (sub-tasks) and each part is trained alone before the complete task is practiced. This approach is called the part or *partial training methodology* [SKI07]. The idea behind is the following: since a complex task is too difficult for a novice to practice, in a first stage the task to train should be simplified. Thus, subjects are trained on separate elements, which are only combined when the trainee has achieved competency for these sub-tasks. The main problem with this approach is that the capabilities required when performing the complete task may be very different from the capabilities acquired when performing each sub-tasks itself. This makes the integration of the acquired sub-task proficiencies into the entire task training problematic (e.g. [CS96]). To resolve these difficulties, Frederiksen and White introduced a training method called Hierarchical Task Decomposition, in which partial tasks (sub-tasks) are created with respect to a detailed task analysis [FW89]. The analysis focuses on the goals and subgoals of the whole task, as well as the knowledge and skills that are required to develop them. Indeed, the hierarchical part task training approach solves most of the integration problems, but it has only limited generalization power to changed task conditions [FBG*89].

Another approach for partial task training is the emphasis change protocol [Gop07], which has been described in Section 3.3.5. Here the parts are not separated from the whole task, but rather at different times one part of the complex task is at the focus of training, what solves the integration problem. During training the emphasized element always varies in relation to all other components, which forces the trainee to pay attention to its performance. Through this attention, the trainee does not loose sight of the relationship and interaction between all sub-tasks [Gop07].

3.4. Measuring of Skill

In order to assess the value of a skill training system, the measuring and analyzing of the trainee's skill level is a crucial factor. Therefore, common methods for the measuring of skills and skill acquisition are described in this section. Since it is not possible to consider all possible measuring techniques, only the more general and widely used methods are considered.

A common method to measure skill acquisition is to analyze behavioral measuring units. Amongst others, those behavioral measures are changes in the average speed and accuracy of responses, as well as the decrease of their variability [FHB03]. Two of the behavioral measures of skill acquisition, namely accuracy and reaction time, are described below. Subsequently also the speed/accuracy trade-off and the learning curve are delineated.

3.4.1. Accuracy of Responding

Accuracy of responding refers to the percentage of responses made by a participant that correspond to a predefined definition of a correct response [SKI07]. The International Standard on ergonomic requirements for office work with visual display (ISO9241-11) states that "accuracy can be measured by the extent to which the quality of the output corresponds to the specified criteria" [ISO9241-11, p19]. That is, accuracy is represented as a percentage of correct responses of the total number of trials. Through practice accuracy can be improved according to a power function. More details about such a power function will be given in Section 3.4.4.

Different ways for recording the accuracy of a task performance exist. Assessing accuracy may be simple (e.g. checking if the subject pressed the correct button), which mainly occurs in laboratory tasks, but may also be more complex in real world tasks. For example, Via et al. [VKGM02] measured the accuracy of an anesthetic task by rating the performance of anesthetist participants as they completed a simulated anesthetic task. For complex tasks, several assessment techniques may be required for an accurate assessment of performance. Another important fact is that in test of skill acquisition appropriate high-level cognitive strategies are assessed [RCC93].

3.4.2. Reaction Time

The time interval between stimulus probes and responses can be analyzed by measuring reaction times. In such measures the combined time of executing all processing, selection, and execution activities are reflected. It has been shown that through practice the reaction times decrease according to a power function already mentioned above, that will be described further in Section 3.4.4. In the context of motor skills acquisition the reaction time interval is usually divided into the two major components *response latency* and *movement time*. Response latency specifies the time interval between the presentation of a stimulus and the initiation of a response movement, i.e. it reflects the response selection processes. Hence, it is sensitive to variables such as time uncertainty, number of alternatives, and response probability. The difficulty and requirement of movement are reflected by movement times. Movement times are sensitive to accuracy, distance force, and type of required movements. Through practice both response time and movement latency decrease [SKI07] (see Section 3.4.4).

Donders [Don69] suggests in this approach to distinguish between simple reaction times (responding to a single probe with a single response) and choice reaction times (selecting the correct response out of several possible responses). He proposes that the mental processes associated with response selection can be measured by subtracting the simple reaction time from the choice reaction time. Simple reaction time tasks, which are easy to perform in laboratory settings, differ largely from real world tasks, which are better analyzed by using choice reaction time tasks [WH99]. The so-called Hicks-Hyman Law suggests that reaction time increases by a constant amount each time the number of choice alternatives doubles [CMN83].

3.4.3. Speed/Accuracy Trade-Off

When humans try to respond more quickly than their current expertise allows, they make more errors. Similarly, when they try to perform more accurately, their performance becomes slower. The mutual relationship between reaction times and errors is reflected by the speed/accuracy trade-off ([Fit54], [WH99], [Woo99]). When instructions are given to participants the relationship between speed and accuracy alters. This relationship may also be influenced by the expectations, motivations, and inclination of participants [FHB03]. These are important facts to be considered for the exploration of performance measures.

The qualitative relation between speed and accuracy has been rewritten in the so-called Fitts' law (e.g. [Fit54], [Mac92]). There are important applications of this law in ergonomics and especially in the design of human computer interfaces ([GBL04], [Mac92]). Using Fitts' law the movement time *(MT)* in a pointing task is related to the size and distance of the target [Fit54]. The law states that the movement time (i.e. time that is necessary to point at a target) increases as a linear function of task difficulty (ID for index of difficulty). The task difficulty *(ID)* represents the amount of information that is necessary to specify the requested spatial accuracy (i.e. a target of width W located at a distance D):

$$MT = a + bID, \text{ with } ID = log_2(\frac{D}{W} + 1) \quad (3.1)$$

with a representing the time lag up to the initiation of the movement, that is the intercept of the start and stop time of the device (e.g. the time required to press a button), and b denoting the slope of the device (i.e. the inherent speed of the device). Since a and b are device- and environment-dependent constants, they need to be empirically determined. D represents the distance between the starting point and the center of the target. The width of the target measured in moving direction (i.e. along the axis of motion) is given by W. The increment by *1* results from a modification of the original form proposed by MacKenzie [Mac92] in order to avoid negative *ID*s for small values of the fraction *D/W*. Thus, especially for small values D/W this form deviates from the original one, but it became apparent that it fits better to measuring data.

3.4.4. The Learning Curve

In Section 3.4.2 it has been described that through practice accuracy and reaction times (on simple and choice reaction time tasks) decrease [WH99]. Those changes in performance over the number of practice trials can be graphically represented by the learning curve (e.g. [NR93], [SKI07]). A power function, which relates the task performance times to the number of practice trials, can be defined. According to this function, task performance times decrease with the number of trials. The power function shown in the Figure 3.4 is represented by the general function, $T_N = b \times N^{-a}$. Here b is a scalar, N is the number of trials or repetitions, -*a* is the slope of the function (rate of learning), and T is the trial completion time or performance level. A similar learning curve can be defined for accuracy. The power law of practice, which describes the relation between reaction time (RT) and number of practice trials, turned out to be an adequate descriptor of performance improvement through practice, including all modes of motor behaviors, perceptual, memory, and decision tasks [NR93]. The power law describes that the logarithm of the time necessary for performing a particular task decreases linearly with the logarithm of the number of practice trials. This means that practice improves performance.

The slowing improvement after the first few trials, improvements over a large number of trials, unstable performance at early stages, and reduced variability with increasing trials are common features of all learning curves ([Log88], [NR93]). This shows that the power law fits to the data of a large number of different skills. Newell and Rosenbloom [NR93] detected a linear relationship between the logarithm of the number of trials and the logarithm of the measured variable (e.g. response time). They used this relationship to predict the logarithm of a trial time T_N, given the logarithm of a trial number N. In the study of skill acquisition both the learning curve and the log transformation are helpful, as they can be used to estimate performance gains over time [SKI07].

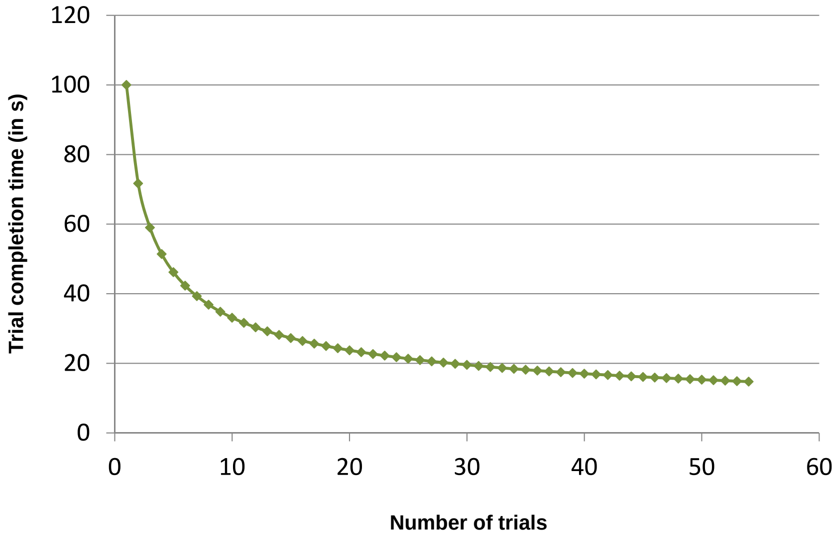


Figure 3.4.: The power law relationship between practice and trial completion time (hypothetical values).

3.5. The Potential of Multimodality

In essence an acquired skill can be considered as multimodal mediated knowledge. Hence, the integration of multimodal interfaces in the skill acquisition and training process provides a promising alternative to the traditional trainer-trainee approach of skill learning [GBK*08]. During the performance of the training task, the trainee can be supported by computer-based multimodal interfaces. The interaction between the trainee and the computer (human-computer interaction) consists of a continuous exchange of task-relevant multi-sensory information. Via such multi-sensory information the task conditions can be adapted in order to improve the trainee's performance (e.g. [GBK*08]).

In daily life it can be observed that the use use of more than one modality can enhance perception and improve performance. So, for example, listening to people in face-to-face conversations is easier than listening in telephone conversations (mainly if the spoken language is not the mother tongue). While in telephone conversations only auditory information is transmitted, face-to-face conversations usually involve also visual information, such as gestures and movements of mouth and lips [GBK*08]. The same applies for noisy conditions: by observing the speaker's lips movements and gestures the lost auditory information can be compensated [SP54]. Furthermore, it has been shown that the reaction time for the detection of visual signals improves if a sound is provided temporally close to the visual stimulus [DS01]. Another approach demonstrated the potential of the combination of auditory and visual information on the example of a pattern judging task [HB69]. The statement that no single sensory signal can provide reliable information about the 3-dimensional structure of the environment in all circumstances has also been explored in several studies ([EB04], [MBS05]).

The human's perceptual-motor abilities result from crossmodal processing of the surroundings. Hence, a coherent perception can evolve from a crossmodal binding of distinct sensory advices with different amounts of reliability [LL04]. Spence [SPD04] showed that auditory stimuli can modulate tactile perception. He observed that by presenting information in one sensory modality, the human's attention can immediately be drawn to other sensory information presented at the same location (e.g. tactile sensations). The humans's process of integration of multi-sensory information to one coherent perception is a multi-stage process. It runs through a cascade of of synergistic processes at different stages of processing [CT04].

Murray [MMM*05] showed that providing additional haptic information to another modality improves the human's performance. Pairs of simultaneously presented haptic and auditory signals were detected faster compared to the unimodal presentation of those signals. Experiments considering texture discrimination tasks conducted in another study led to the same result [Hel82]. In the study of Lederman et al. [LMHK02] it was observed that the user's confidence level regarding his own judgments is higher when multimodal information is provided. Also an interesting finding is that the reaction time to simultaneous visual and tactile stimuli is faster than the reaction time to simultaneous dual visual or dual tactile stimuli [FCAB02]. If additionally auditory stimuli are added, the detection of those tri-modal is even faster compared to the shortest reaction time of the bi-modal combinations [DC04].

To summarize, according to the nature of the task and the robustness of sensory input, the influence of the modalities may vary. By using multiple modalities a richer representation of the world can be created and thus the user's perception can be enhanced. Besides, perceptual-motor activities can be influenced and attention processes can be modulated. Based on these findings it can be assumed that the presentation of information using multiple modalities during training can improve and accelerate the user's understanding of the training task and the actions to perform.

3.6. Summary

In this chapter it has been shown that a number of factors must be considered for the development of efficient skill-training systems. So, for example, it is important to decompose complex skills into their sensorimotor and cognitive sub-skills in order to identify exactly which skills must be trained. Furthermore, it is relevant to understand the learning process of humans and to support it using training methodologies that have proven their worth. In addition, it has been pointed out that providing information using multiple modalities can enhance the user's perception and hence, can potentially improve training.

A further important factor for the development of training systems is the use of proper technologies to realize and apply training methodologies, to present information and feedback in such a way that it provides the best support, and to gather information that is necessary to provide adequate feedback. Such technologies will be described in the following chapter.

4. Technologies for AR-Based Skill Training

In this chapter, important technologies for the development of multimodal Augmented Reality-based training systems are described. This includes a short introduction to Augmented Reality, a discussion of Augmented Reality display devices, and the description of relevant capturing and rendering technologies.

4.1. Augmented Reality

Augmented Reality (AR) deals with the combination of real world images and virtual objects. Computer-generated virtual objects are overlaid into the user's field of view to provide additional information (i.e. to enhance reality). Thus, in contrast to Virtual Reality, where the user is totally immersed in the artificial environment, in Augmented Reality the real environment is still perceived. Azuma's [Azu97] definition of Augmented Reality states that (1) Augmented Reality combines the virtual and the real world, (2) is interactive in real-time and (3) is registered in 3D, that is, the registration between real and virtual data has to be performed in real-time. In this point Augmented Reality systems differ from technologies, in which real and virtual data are synthesized off-line (e.g. in the movie post-production). In Augmented Reality systems the alignment of virtual objects to real images is accomplished in real-time.

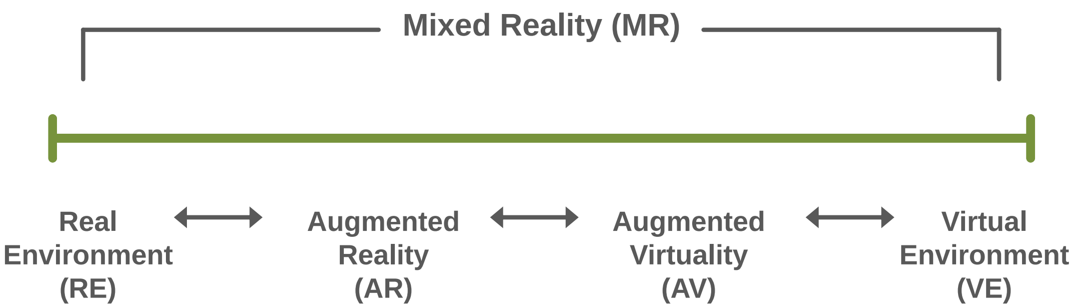


Figure 4.1.: Milgram's Mixed Reality Continuum (adapted from [MK94]).

Milgram [MK94] defined a Mixed Reality Continuum which describes a continuous transition from the real environment to a pure virtual one (see Figure 4.1). During this transition the continuum passes through Augmented Reality (closer to the real environment) and Augmented Virtuality. While in Augmented Reality the reality is augmented with virtual objects, in Augmented Virtuality the virtual world is augmented by real objects.

To align the virtual objects with the real world, a common approach is to estimate the position and viewing direction of the camera by analyzing the digital camera image, which is still a challenging field in unprepared environments and under real world conditions. To display the resulting augmented images, display devices like Head-mounted Displays (HMDs), handhelds (e.g. ultra-mobile PCs), or tablet PCs are widely used. Since existing head-mounted devices are not ergonomic and comfortable to use, many Augmented Reality applications have been developed on handhelds and tablet PCs.

Augmented Reality has shown its potential in various indoor application areas such as industrial maintenance and surgery, and outdoor application fields like architecture, tourism and entertainment. For example, a city skyline can be augmented with virtual 3D models of buildings or virtual reconstructions of an ancient building can be placed on the ruin in order to show, how it appeared many years ago. Another application field are TV live broadcasts, where Augmented Reality can be used to superimpose additional information for the viewer in order to clarify situations, like for instance an offside situation in a football match.

4.2. Display Devices for Augmented Reality

The utilization of appropriate devices is an important factor for the usability of a training system. This includes the handling of the involved devices as well as the interaction with them. The usability of Augmented Reality-based systems is mainly determined by the utilized display devices. In the following different Augmented Reality display devices are described and analyzed.

4.2.1. Optical See-Through and Video See-Through Displays

In general, Augmented Reality displays can be divided into two categories: *video see-through* displays through which the real world is observed in a video image (see Figure 4.2), and *optical see-through* displays, which consist of a semi-transparent screen that provides a direct view to the real world (see Figure 4.3). Both kinds of displays have advantages and disadvantages.

The technique used by video-see through displays is called video mixing. Video mixing merges live video streams with computer generated graphics. The result is displayed on a screen. Hence, the user cannot observe the real world directly, but through the image projected on the screen. Usually a camera mounted on the head or display delivers a live video stream that provides the view to the real world. The main drawbacks of video-see through devices are the limited field of view that results from limitations of the applied optics, and a lack of resolution due to limitations of the applied (miniature) displays.

Optical see-through displays make use of optical combiners, which displays computer graphics within the real environment or the user's visual field. Thus, optical-see through devices can offer a superior view to the real world compared to video-see through displays. The user has an unmodified

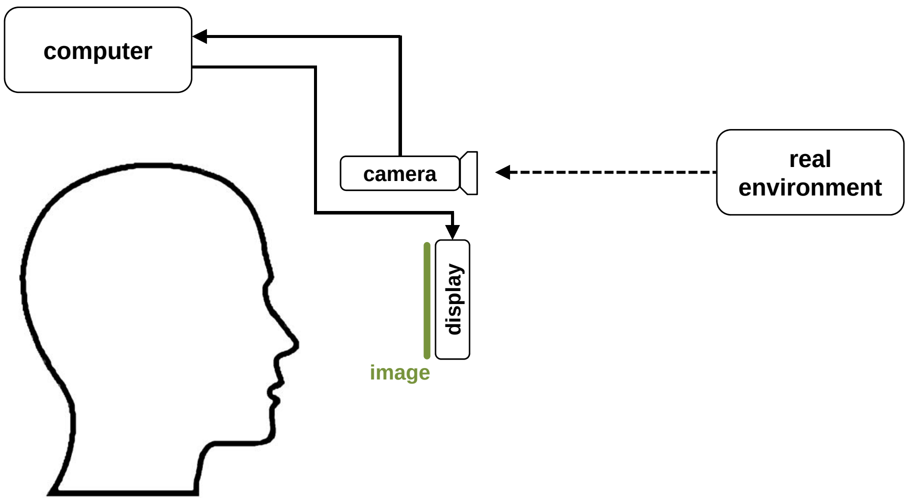


Figure 4.2.: Abstract video see-through concept (adapted from [BR05]).

view to the real environment which can be observed in full resolution and without any delays. Only the augmented graphics suffer from the resolution of the displays.

A major disadvantage of optical see-through devices is that they require a difficult, user- and session-dependent calibration and a precise estimation of the camera (e.g. an accurate tracking of the head) to ensure an exact graphical overlay, while for video-see through displays the graphics can be integrated on a pixel-based basis. Furthermore, the display of the computer generated virtual objects is delayed, because of the computing time that is required for image processing, pose estimation, and integration of the graphics. When the camera (mounted on the device or on the user's head) moves, a nasty swimming effect can occur. Through video-based devices the graphical objects can be drawn at the right time, since real environment (through video image) and virtual objects are displayed to the user at exactly the same time.

4.2.2. Head-Attached Displays

Head-attached displays are devices which the user is wearing on his head. Hence, the user has both hands for interaction and can move freely, what is an advantage compared to hand-held devices. The most commonly used types are retinal displays and head-mounted displays.

Retinal displays use low-power lasers to project images directly onto the retina of the human eye ([PFV98], [Lew04]). This results in a much brighter and higher resolution image and a wider field of view than a screen-based display can produce. Also the level contrast in the image is higher. Since retinal displays consume only low power, they are well suited for mobile outdoor applications. Future retinal displays may also provide dynamic re-focus, extremely high resolution, and full-color stereoscopic images. Retinal displays have also disadvantages, such as a fixed focus length that results from

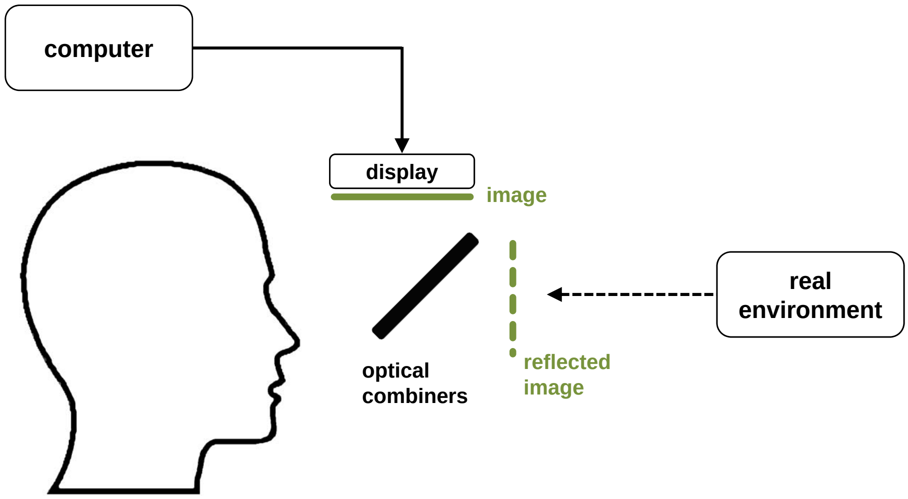


Figure 4.3.: Abstract optical see-through concept (adapted from [BR05]).

the complete by-passing of the ocular motor system by projecting directly onto the retina, or the lack of good stereoscopic versions.

Two different types of head-mounted displays exist: optical-see through and video-see through devices. Optical see-through HMDs usually consist of transparent LCD displays or half-silvered mirrors and make use of the optical combination technology, while video-see though HMDs use video-mixing to display augmented images within a closed-view head-mounted miniature display (cf. Chapter 4.2.1). A detailed comparison between optical-based and video-based HMDs is given in [RHF94]. Since HMDs are worn on the head and can be connected with a mobile computer that the user can carry for instance in a backpack, they can be used for mobile applications.

Even if HMDs are widely used for Augmented Reality applications, several drawbacks can be noted: As mentioned in the previous section, the resolution provided by HMDs (mainly video-based HMDs) is too poor to allow for arbitrary augmented objects. For example, textual information may be difficult to read for the user. Besides, the optics integrated in a HMDs are typically heavy. This results in unergonomic and uncomfortable devices or ergonomic devices with a low image resolution. A problem occurs during fast head-movements which may create a feeling of discomfort and sickness of the user due to the head-attached image plane [PCS*00]. Another problem of optical-see through HMDs, called fixed focus length problem, is caused by the inconsistent image depth between the head-attached image plane and the real environment. Objects within the real environment are sensed at a different depth than the displayed virtual objects. Hence, the user's eyes either have to continuously shift focus between distinct depth levels, or one depth level is perceived as blurred.

4.2.3. Hand-held Devices

As described in the previous section, head-attached displays suffer from usability and registration difficulties. To avoid those problems, alternatively hand-held display devices can be used. Hand-held displays such as tablet PCs, ultra-mobile PCs, PDAs, and, more recently, smartphones work within arm's reach. Those devices combine display, processor, memory, and interaction interface in one single device. Thus, the cable tangle that usually occurs when using head-attached devices is avoided. Hand-held displays are typically equipped with a screen and a camera (on-board or externally attached) and work in video-see through mode (like the preview window of a digital camera). Before the captured live video stream is displayed on the screen, it is overlaid by virtual objects.

Although rarely used, there are also some examples for optical-see through hand-held devices. The one introduced by Stetten et al. [SCHB01] consists of an ultrasound transducer scanning ultrasound slices in front of it. A small flat-panel screen displays sequentially the ultrasound slices which then are reflected by a planar half-silvered mirror in alignment with the visual outer surface of a patient.

However, more common are video-see through hand-held displays. The probably first one was introduced by Rekimoto [Rek95]. He presented a hand-held palmtop consisting of a hand-held LCD screen and a small camera mounted on this screen. To increase the processing power, the hand-held device is connect to a computer and a video compositing engine. Using bar-codes, which are attached to real objects, virtual information is overlaid when the bar-codes are recognized in the video image. A user study showed that subjects using this system were able to "visit" three bar-codes in less than half of the time required when using a head-mounted display. Besides, the subjects perceived the hand-held device as more ergonomic than the HMD. They only found fault with the weight of the device and with not being able to operate with hands free.

Since the processing power of today's hand-held devices has increased significantly, for many AR applications the outsourcing of computational processes to external computers is no longer required. By using tablet PCs, which are the most powerful hand-held devices, a wide range of Augmented Reality applications can be supported (see Figure 4.4). Beside the hand-held property, they also offer an intuitive touch (pen- or finger-input) interface and considerable computing power. Furthermore, it provides a much brighter and higher-resolution display than HMDs and thus allows for full-color overlays. Tablet PCs equipped with a dedicated graphics board do actually allow for applying advanced shading techniques and rendering of high-polygon 3D models.

Several Augmented Reality applications have been implemented on tablet PCs, such as the archaeological application presented in [VIK*02]. In this work a tablet PC is used to display virtual reconstructions of ancient buildings as overlays onto their original position (e.g. onto a ruin). In another example, a tablet PC acts as an AR-based shopping assistant that provides information about the shops linked to the corresponding locations.

If the processing power of the targeted hand-held device is still not sufficient (e.g. when using PDAs or smartphones), wireless networking can be used to outsource computations. In this case, the hand-held is treated like a client, while images are transmitted and processed to external workstations. The

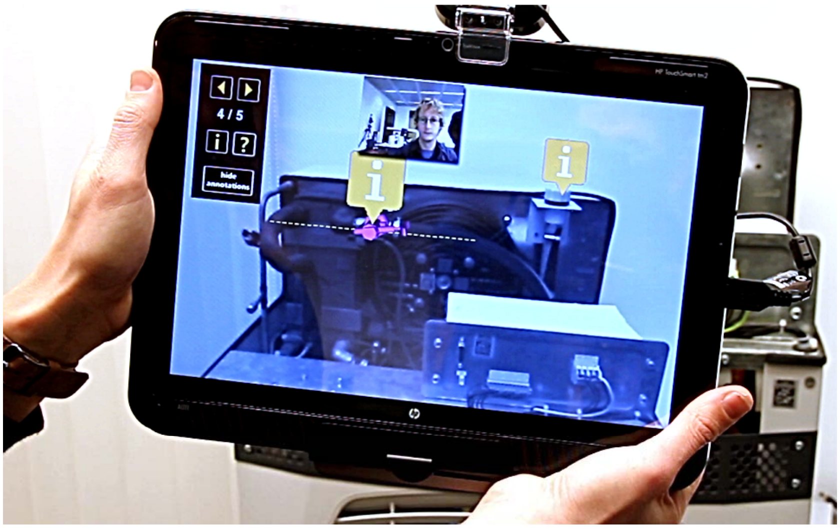


Figure 4.4.: Augmented Reality on a tablet PC.

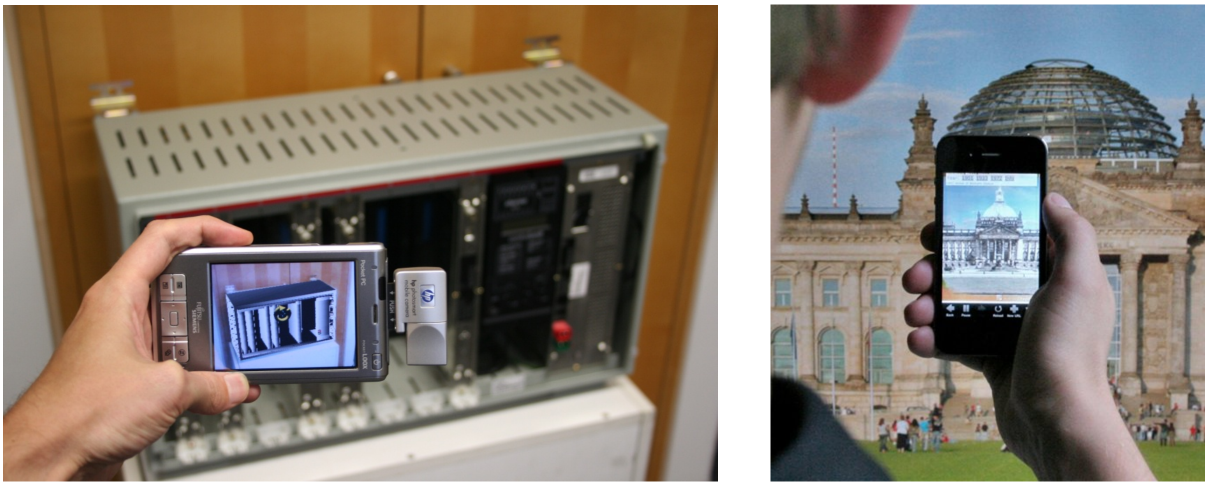
augmented image is sent back to the hand-held where it is displayed on the screen (e.g. [GKRS01], [GFM*03]).

PDAs and smartphones still cannot keep up with tablet PCs in terms of processing power, but are a reasonable alternative for simpler AR applications. Moreover, such devices usually provide a built-in wireless connectivity. Examples for Augmented Reality applications on PDAs are given in [MLB04], [HBO05], [Hec07], and [Wag07]. Figure 4.5 gives an impression of Augmented Reality running on PDAs and smartphones.

4.2.4. Spatial Displays

Spatial Augmented Reality displays are completely detached from the user. They can be divided into three different categories: *video-see through*, *optical-see through* and *projection-based* displays.

Screen-based video-see through displays make use of the video-mixing technologies (see section 4.2.1). The augmented images are displayed on regular monitors. Feiner et al. described those screen-based systems as "window to the world" [FMHS93]. The amount of the user's field of view that can be augmented with virtual objects (also called immersion) depends on the size of the monitor, its distance to the user, and its spatial alignment relative to the user [BR05]. The resolution of the displayed image is restricted due to the resolution of the video stream of the real environment [RHF94]. A screen-based AR system does not require any special hardware in addition to a standard PC equipment. If mobility or optical-see through characteristics are not required, screen-based video-see through systems are the most commonly used approach.



(a) (b)

Figure 4.5.: Augmented Reality on (a) PDA and (b) smartphone. (Fraunhofer IGD, 2007/2010)

In spatial optical see-through systems virtual objects are aligned within the real physical environment by using optical combiners (see section 4.2.1). This technique causes some problems: The user's interaction with real objects during observing the augmentations might be difficult, since those objects are located behind the screen. Furthermore, virtual objects outside the screen area are cut. Like screen-based systems they cannot be used for mobile AR applications. The maximum number of concurrent system users depends on the utilized optics.

Compared to video-see through systems, spatial optical-see through installations provide a higher display resolution (only the resolution of the virtual objects is restricted) and a larger field of view. In comparison with head-mounted optical-see through devices the calibration of the spatial systems and the accommodation of the user's eye to the visualization is much easier. In addition, spatial systems are more ergonomic to use. Since spatial optical see-through systems are only used for non-mobile indoor applications, the environment is much more controllable.

Projection-based spatial displays project images directly on the surface of physical objects and not in the user's field of view ([UUI99], [BGW*02]). Multiple projectors can be used to increase the display area and the quality of the projected images ([RWLB01], [BCK*05]). Projection-based displays are also more ergonomic than head-mounted systems and provide a much larger field of view for the user (theoretically unrestricted). Since the graphical objects are rendered near their location in the real environment, the user's eye can better accommodate to the visualization. When using multiple projectors, the geometric and color calibration of the projectors may be complex and can cause difficulties. Furthermore, the display area is restricted to the size of the projection surface. Another problem can occur if the user casts shadows when he is interacting.

4.3. Capturing for AR-Based Skill Training

In order to provide appropriate feedback to the user and to respond to his actions during training, various kinds of information need be captured. These factors comprise human body motions, applied forces, objects used for interaction, cognitive factors, and much more. The number of aspects involved in the user's performance of a complex skill is enormous, and it is impossible to consider all these factors in the training system. Hence, the aspects that are most relevant for providing adequate feedback during training have to be chosen and captured. However, the training system should be designed to allow for the integration of many different capturing technologies and the handling of the corresponding captured data, in order to be adaptable to different kinds of skill training (e.g. training of maintenance, rowing or surgery skills). Following, some of these capturing technologies are described.

4.3.1. Tracking for Augmented Reality

In order to determine position and orientation of the virtual overlays in Augmented Reality applications, the camera position and viewing direction (camera pose) must be tracked. There are many approaches using different technologies for detecting the camera pose. Those approaches comprise non-visual tracking technologies, such as magnetic [TLN01] or ultrasound [SD05], as well as visual tracking technologies which analyze the video stream delivered by at least one camera in order to estimate the camera pose (e.g. the user's head in case of using a HMD). Since most of the AR display devices (see section 4.2) are already equipped with a camera and the processing power of mobile devices is yet respectable (and continuously increasing), visual tracking is becoming more and more appealing as a low-cost registration technology for Augmented Reality. There has been a lot of research in the field of visual tracking, but especially tracking in unprepared (no markers) and unknown scenes is still a challenging field. The major approaches for detecting the camera pose are *inside-out* and *outside-in* tracking.

4.3.1.1. Inside-Out Tracking

Typically for Augmented Reality a camera is mounted on the user's head or the display device, and this camera's position and orientation is calculated relative to detected features in the environment. This technique, when the camera is located inside the space to track and features in the environment are analyzed, is called inside-out tracking.

Various approaches concerning this technology exist. One attempt is to place artificial markers, so-called *fiducials*, with special geometric or color properties in the scene. Those special properties facilitate the extraction and identification of the markers in the video frame. The position of the markers in the world is known and hence the camera pose can be estimated relative to the detected markers.

A very popular library for tracking using passive fiducials is the ARToolkit which is based on the work of Kato et al. [KB99]. The markers consists of heavy black square outlines into which a unique pattern is printed. The markers are detected by applying image thresholding, line detection, and region extraction technologies. The region to extract is characterized by a boundary of four straight lines. The identification of the detected regions is realized by matching them against templates. Each marker leads to a full 3D pose.

The Intersense VisTracker [FN03] is based on an approach using circular fiducials to calculate the camera pose, which has been presented by Naimark et al. [NF02]. This marker consists of fifteen binary cells which are arranged inside a circle. The authors implemented a self-tracking camera-sensor unit in hardware. This unit contains algorithms for detecting and identifying the markers. A sensor fusion filter combines information from previous tracked frames and inertial sensors to reduce the tracking workload, what is necessary since the extraction and identification of the markers cannot be processed at full frame-rate on the used hardware.

Fiala presented a design approach for highly reliable fiducial marker systems [Fia10]. He states that marker systems should provide reliable information with unknown printing methods and cameras. The reliable systems should also be able to work under uncontrolled lighting conditions without detecting markers falsely or reporting the wrong marker ID. Furthermore, he developed the ARTag fiducial marker system [Fia05] which combines an edge-based detection method that can cope with inconsistent lighting conditions and partial occlusions with a reliable digital coding system used for the identification and verification of the markers.

A lot of research has also been done in the field of markerless camera tracking. Wuest et al. presented an adaptive model-based, or rather edge-based, line tracking approach, in which lines are detected by continuously learning the visual properties of edges after a successful camera pose estimation [WVS05]. This method requires the 3D model (VRML) of the object to track, which is used to generate a 3D line model with respect to the camera pose estimate by performing a visibility test on the standard graphics hardware. For every sample point of the projected 3D model lines a search for gradient maxima in the image is then conducted. By minimizing the distances between the projection of all sample points of the 3D line model and the most likely matches found in the image the camera pose is updated (see Figure 4.6).

A point-based approach for markerless tracking was presented by Bleser et al. Only for the initialization minimal 3D knowledge in the form of a 3D model of one scene object is required. Similar to the approach of Wuest et al. described above, at first a line model of the rendered CAD model is created. For calculating the initial pose, this line model is registered with the video image. Then, during tracking the scene is reconstructed automatically and hence the camera is not restricted to the part of the scene which is covered in the CAD model. This is done by detecting point features in the video images and tracking these features from frame to frame using a template matching algorithm. For the rough determination of the initial 3D feature locations, linear triangulation is applied. These locations are then refined by using the Extended Kalman Filter framework. Based on the refined locations the camera pose is estimated.

Rosten et. al describe a markerless tracking method that combines edge and point tracking [RD05]. Using the FAST algorithm corner points are extracted from an offline model that is restricted to edges

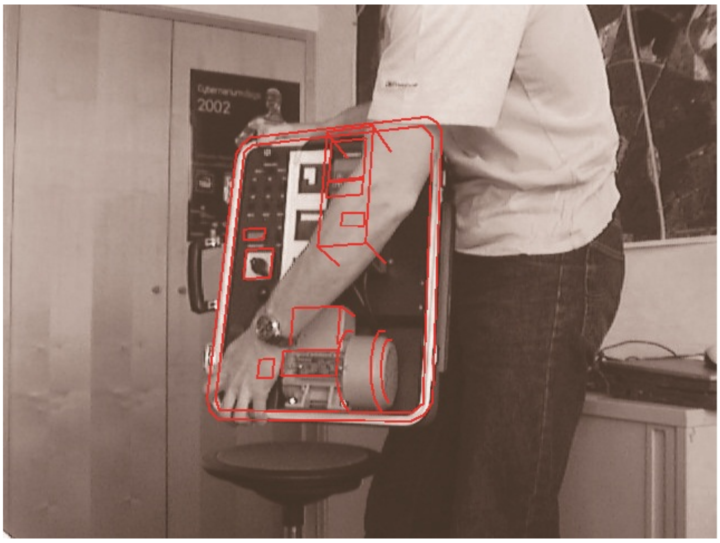


Figure 4.6.: Line tracking example [WVS05].

and are matched from the previous to the current frame. That way, a motion estimate between two subsequent frames is provided which can exploit clutter. Amongst others the authors describe a non-linear method for combining the two tracking systems (line-based and point-based).

A further approach for estimating the camera pose in unknown scenes, called PTAM (parallel tracking and mapping), has been presented by Klein and Murray [KM07]. In this approach, tracking and mapping are split into two separate tasks and processed in parallel threads. Here, one thread handles the tracking of a hand-held camera by receiving images from the camera and maintaining a real-time estimate of the camera pose. The other thread generates a 3D point feature map of previously observed video frames by continually refining a feature map initially created using a stereo technique. The tracking quality provided by this method is appropriate for small Augmented Reality workspaces. In subsequent work this method has been adapted to run even on camera phones [KM09].

4.3.1.2. Outside-In Tracking

In outside-in tracking the imaging sensor is mounted outside the space to track and observes the tracking area. Especially when multiple cameras are used, outside-in tracking can deliver very accurate results concerning the position, but compared to inside-out systems it can not generally produce the same accuracy for orientation. Besides, the range of an outside-in system is restricted by the range of the camera mounted in the environment. By using multiple cameras this range can be enlarged.

A typical example for outside-in tracking is the use of active fiducial markers, such as infrared LEDs. When an appropriate filter is used at the imaging sensor, only the fiducials are visible to the sensor and all ambient light is virtually eliminated. Thus, the difficulty and workload for the tracking can be reduced. When those LEDs are mounted on the user's head or the display device, its pose can be tracked by applying outside-in tracking. One early example is given by Ferrin [Fer91], who used LEDs mounted on the helmet of a pilot to track the pose of the pilot's head.

Klein et al. [KD04] describe a tracking approach for tablet PC-based Augmented Reality, where an outside-in tracker observes the tablet PC equipped with LEDs to generate robust pose estimates with low-accuracy, while an inside-out tracker which is running on the tablet PC works on the video stream of the tablet-mounted camera. The inside-out system tracks natural features in the environment and is able to produce high accuracy pose estimates, which are then combined with the low-accuracy estimates of the outside-in tracker by using an Extended Kalman Filter. Thus, robust and accurate Augmented Reality on a tablet PC can be achieved.

4.3.2. Hand Tracking

The human operates a lot with his hands. Especially when performing maintenance tasks, the user operates mainly with his hands. Hence, the tracking of the user's hands is an interesting aspect in the context of training of maintenance skills.

Different approaches for tracking hands exist. One approach that uses optical tracking is presented by Malerczyk [Mal08]. The user is observed by two cameras in front of a large displaying screen while he is performing gestures. The segmentation of the user's hand is performed on 2D image basis. The video images captured by the stereo system are smoothed using a Gaussian filter mask, and edge images are calculated using a standard Sobel kernel. These edge images are then compared with edge reference images of the empty interaction volume, which have been taken on system startup, by calculating difference images. The resulting image pairs are then segmented at a predefined threshold. Since the cameras are positioned at the left and at the right side of the user, and since the user is always interacting with a screen in front of him, the lowest extracted segment larger than an adequate segment size is chosen as the user's hand. Then an approximately square box at the bottom of the segment is selected in order to get a segment only containing the hand, since often also the users forearm is extracted into a segment. This box segment containing the hand is used for feature extraction. The position of the user's hand in the world coordinate system is determined by projecting the centers of gravity in both images into 3D space.

A further example is the work of Sudderth et al. [SMFW04] in which probabilistic methods for visual tracking of a 3D geometric hand model from monocular image sequences are described. They show that the geometric models commonly used for hand tracking naturally have a graphical structure which can be utilized to track the hand's motion using nonparametric belief propagation (NBP) ([SIFW03], [SII*10]).

Petersen et al. [PS09] present a method for real-time hand detection and tracking by focusing on posture invariant local constraints, which exist on finger appearances, instead of considering the hand as a whole. This approach is based on selecting local regions that comply with a number of geometric and photometric posture invariants. Using experimental test the authors showed that even if low quality skin color information is provided, this method works robustly above cluttered background. Thus, no calibration to the user's skin color is required before use.

Other approaches for capturing hand motion use electro-mechanical or magnetic sensing devices like data gloves (e.g. [Fox02]). These devices can measure the location of the hand and the finger joint angles. They provide an application-independent set of real-time measurements of the whole hand. But since those devices are worn on the user's hand, they impair a natural hand motion and handicap the user's interaction with an object. Furthermore, they require complex calibration and setup procedures to achieve precise measurements, and are very expensive.

4.3.3. Motion Capturing

Motion capturing (mocap) refers to the digital recording of the movements. Mostly, markers are applied to dynamic subjects, for example human faces, bodies, or robot arms, in order to track their motions. A common motion capture approach is to use optical tracking technologies. Traditional optical motion capture systems have two elements: One is the equipment that is attached to the subject and transmits the motion information. The other one is the equipment that is arranged around the subject and receives the information. The equipment attached to the subject usually consists of retro-reflective spheres. Cameras, which are arranged around the subject, have a light source co-located with the camera lens to illuminate the retro-reflective markers. Hence, in the camera images the markers appear much brighter than the background under controlled lighting conditions. Those optical mocap systems usually use multiple cameras to triangulate the 3D position of markers [SKI09]. A shortcoming of such systems is that they have to operate in low-light. Furthermore they are very expensive. An example for an optical motion capture system using retro-reflective marker spheres can be found in [KHYN02].

Tanie et al. [TYN05] enhanced the traditional approach by using a suit covered with a retro-reflective mesh instead of using marker spheres [TYN05]. That way, they can attach a huge number of markers on the subject. The mesh is made of retro-reflective tape, what allows for using the cameras of traditional optical motion capture systems. The connectivity information provided by the mesh is exploited to improve the efficiency and accuracy of the reconstruction process, and hence a fast and precise measurement of the markers is achieved.

There are also approaches for markerless motion capturing. Yoon et al. [YMG09] present a system for markerless real-time 3D motion capturing to facilitate the analysis of three-dimensional human body movements during a sprint start. The system uses multiple cameras. In a first stage an adequate segmentation of video streams is carried out. Then follows a robust background and silhouette extraction, what enables a subsequent 3D real-time reconstruction of the target object's volume. In order to estimate the 3D pose of the body, a reduced articulated human body model that complies with the HAnim standard [ISO_IEC_FCD_19774] is fitted into the 3D reconstructed volume model. The pose estimation is realized by deploying Pseudo-Zernike moments (PZM).

Unzuneta et al. [UPBS08] present a method to reconstruct human's full-body movements based on the positions of the human's end-effectors in order to make it usable within low-cost motion capturing systems that would only need to track the end-effectors instead of tracking the whole body. They propose an analytic-iterative inverse kinematics method to reconstruct human's full-body movements in real-time. Only the positions of wrists, ankles, head, and pelvis are used as input data for the reconstruction. These points are considered as the principal markers for general full-body movements, i.e. the points that retain the essential information of full-body movements. From these captured data the pose of the humanoid that represents the user is reconstructed using Inverse Kinematics.

The first commercially available markerless motion capture system, called OM Stage, is presented by Organic Motion [Org11]. OM Stage is a video-based motion tracking system consisting of four components. One of them is customizable truss which is covered with a reflective white cloth and thus defines the scan space. Fourteen monochrome cameras which capture video frames of the scan space, or rather of the user to be tracked in the scan space, form the second component. The third component handles the 3D tracking. It consist of a vision processor that generates real-time full 3D motion data from the information delivered from the cameras. The last component a software development kit that streams the real-time 3D data into other software modules (e.g. animation or biomechanics software) to be further processed and interpreted.

Beside optical capturing technologies, also non-optical methods can be used for capturing human movements. For example, miniature inertial markers can be attached to the human's body which transmit motion data wirelessly to a computer with which the motion can be recorded, analyzed, or viewed (e.g. [MJKM04]). With the number of attached inertial sensors also the quality of the movement that can be reconstructed increases. Inertial motion capture systems provide a real-time capturing of full 6D body motion of a human in large capture volumes, but deliver a much lower positional accuracy compared to the accuracy that can be achieved by optical systems.

A further example for non-optical motion capture systems are magnetic systems which use magnetic transmitters and receivers to track the human's movements (e.g. [YKY*00]). Position and orientation are calculate by the relative magnetic flux of three orthogonal coils on the transmitter and on each receiver. Inertial sensors can provide 6D information about the object to which they are attached. The inertial markers (the receivers) are not occluded by nonmetallic objects, but the transmission can be impaired by metal objects in the environment, such as monitors, computers, or cables. A problem is the capture volume which is much smaller than the capture area of optical motion capture systems.

4.3.4. Force Capturing

Beside hand tracking and motion capturing, the recording of forces applied by the human is another important aspect for capturing humans performing sensorimotor skills (i.e. performing the corresponding tasks). One factor of interest is the 6D-force that occurs at a specific position during user interaction. In this sense the term *force* also includes torques. 6D-force means three translational force values directed along the three Cartesian axes and three torque values directed around those axes. One possibility to capture 6D-force is the use of a force-torque sensor. Such a force-torque measuring device is shown in Figure 4.7 [TFR*10]. The device contains a series of force sensors and a microcontroller and is able to provide measures at 100Hz. It is connected via Bluetooth to the main system CPU. This device can be used for example for measuring the force and torque the user applies when loosening a screw.

Figure 4.7.: Force-torque sensor developed by PERCRO (see also [TFR*10]).

A further relevant factor for capturing sensorimotor skills is the capturing of a force profile, since oftentimes the information how force is distributed over the contact surface is important. Therefore, the force to be measured is a vectorial or 3D-force applied to a surface and may be uniformly applied or distributed through several surface elements. The force may also involve friction. The surface itself is characterized by its mechanical impedance that indicates the displacement velocity of the surface according to the applied force. The surface can be for example hard or soft, flat or curved, cold or warm. The measuring of a force profile is strongly correlated with tactile sensation [SKI09]. That is to say, pressure is only one part of a rich tactile information system that also involves surface properties like shape, hardness, temperature and more. Hence, the force measuring device for capturing the force profile should not only measure, continuously or on request, the applied force, but should also be able to measure surface properties (as described above) [SKI09].

One example for an force profile capturing device is presented by Hoshi et al. [HOMS06]. They developed a kind of artificial skin for normal force sensing. It consists of a flexible film containing a network of pressure sensors. The film, that can cover the hand, acts as an artificial skin. The pressure is measured by capacitive effect. The artificial skin also contains a network of capacitive sensors, where each sensor consists of two soft film capacitors lying upon another. Both capacitors vary differently according to the applied stress. This characteristic allows for a simultaneous measuring of the applied normal force and of the properties of the surface over which the force is applied.

A further example is given by Yamada et al. [YMY01], who developed an artificial elastic robot finger skin for controlling grasp forces (grip and lift) when weight and frictional coefficient of the grasped object are unknown. Ridges at the surface of the artificial finger skin imitate the ridges of a human finger. Thus, the "stick" and "slip" effect that occurs when human fingers grasp an object can be emulated (the center of contact area generally sticks, while at the edges the contact condition changes from stick to slip). Similar to the receptors of human fingers or rather fingertips, a pair of tactile sensors is embedded for ridge. To enable that contact force distributes, the surface of the whole artificial finger is curved. That way, the incipient slippage of the ridge that occurs near the edge of contact area can be detected.

4.4. Rendering for AR-Based Skill Training

The term rendering refers to the process of generating a perceptible presentation of data stored in the computer, comprising visual, haptic, and audio information. As described in Chapter 3.5 the application of multimodal feedback can improve training. Hence, technologies for rendering various kinds of feedback, such as visual, haptic, or audio feedback, are needed. These technologies are described below.

4.4.1. Rendering of Visual Information

Since a human absorbs nearly 80% of information through the eyes, the rendering of visual feedback is a very important factor for training systems. The data to visualize can contain various kinds of information like for example 3D models, text, still images, or videos. Hence, the visual rendering module of a skill training system should be able to visualize different kinds of data in real-time.

In general, the rendering of visual feedback passes through three functional stages [BC03]. The first stage is the application stage, in which all stored geometry data and the input delivered by user interaction (e.g. mouse movements and clicks, touch input on a tablet PC, data from capturing systems) or capturing technologies (e.g. position and viewing direction of the user) is read. According to the input, the virtual camera, that is the view to the simulation, or the position and orientation of virtual objects can be changed. The resulting data is passed to the geometry stage, which consists of model transformations (translation, rotation, scaling, etc.), lighting computations, scene projection clipping, and mapping. The lighting computations are important for giving the virtual object a more realistic appearance. They serve to calculate the surface colors of the graphical objects under consideration of the type and number of simulated light sources in the scene, the lighting model, the surface material properties, and atmospheric effects (e.g. fog), what results in a shaded scene. In the last stage, the rasterizing stage, the vertex information delivered by the geometry stage (color, texture, etc.) is converted into pixel information.

In the field of Augmented Reality the realism of the visual augmentations may be an important issue, particularly if the overlays consist of 3D models or animations. In order the user can spatially assess the overlays, his depth perception of the virtual objects is important. One approach to enhance the appearance of the virtual overlays is to consider the real environment, such as real light sources, for the shading of the objects.

A common technique to render objects under real world lighting conditions is called image based lighting (IBL) ([Deb98], [Deb06]). Instead of manually placing direct light sources, this method uses spherical images to capture incident lighting conditions and map them onto the object (see Figure 4.8). This process has been customarily known as reflection mapping and is used to simulate highly reflective surface materials. A generalized version is known as irradiance reflection mapping, where filtered versions of reflection maps are used to simulate other material functions, for instance blurry images for diffuse reflections. In image based lighting, these techniques are used in combination with high dynamic range (HDR) images. Since usual images cannot represent the full dynamic range of incident light, HDR images are required to simulate lighting without the constraint of working only under fixed assumptions about the surrounding lighting condition for which certain material reflections might not look visually pleasing. Thus, image based lighting is not only the most straight-forward method of dealing with global illumination-like effects in Augmented and Mixed Reality applications, but also the method that avoids otherwise complex problems as identifying and tracking direct and indirect light sources in images [SS06].

Figure 4.8.: Image based lighting applied to an augmented 3D model. (Fraunhofer IGD, 2010)

A further important issue is the handling of occlusions between real and virtual objects. When the user interacts, he may occlude (partially) objects on which 3D models are superimposed (e.g. the user's hand may partially occlude a plug which he should remove). If the corresponding virtual object is not occluded by the user, it is difficult for him to spatially assess this virtual object and hence to identify which object is targeted (e.g. which plug he has to remove). This is a critical problem which has not yet been completely solved in terms of accuracy. One approach for capturing depth information of real objects, what is essential for handling occlusions, is to use time-of-flight (TOF) cameras [HA09]. Usually those cameras are quite expensive and do not deliver satisfying accuracy results due to their low resolution. A relatively new device than can be used to capture depth information about real objects is the Microsoft Kinect, which is a combination of a RGB camera, a depth camera (both with a resolution of 640×480 pixels), and a 3D microphone [Mic10]. First attempts have shown the potential of using this low-cost device for capturing depth information for Augmented Reality applications.

4.4.2. Rendering of Haptic Information

As already mentioned, haptic feedback can support the user in acquiring a skill. Therefor haptic rendering technologies are needed to render haptic information presented to the user using haptic devices.

Similar to visual rendering, also when rendering haptic feedback a pipeline is passed [BC03]. According to Popescu [Pop01] the pipeline can consist of three functional stages. In the first stage (collision detection stage), the physical characteristics of the 3D objects (e.g. compliance, smoothness, weight, etc.) are loaded from the database and collision detection techniques are applied to achieve information about colliding objects. The resulting colliding structures are fed to the second stage (force computation stage) which calculates the collision forces under the consideration of the physical simulation model to be applied (e.g. Hooke's law). This second stage handles also force smoothing (adjustment of the force vector direction to avoid sharp transitions between polygonal surfaces) and force mapping (mapping of the computed force to the characteristics of the haptic device). In the last stage (tactile computation stage) the touch feedback of the simulation is rendered and the computed effects (e.g. vibrations, surface temperature) are added to the force vector. The result is sent to the haptic device.

There are two different categories of haptic feedback: One is the kinesthetic feedback which, for example, transmits information about applied forces and torques or guides movements. The other one is the tactile feedback. This kind of feedback corresponds to a mechanical interaction with the user's skin and can be presented, for instance, in the form of vibration or pressure (e.g. [SEWP10]).

However, in the context of Augmented Reality-based training the user interacts mainly with real objects and the haptic feedback channel is rather used to guide the user than providing him contact information about the virtual objects. Consequently, the forces or effects (like vibrations) presented to the user do not necessarily result from the behavior of the virtual objects. Such effects and forces may be predefined patterns or calculated during runtime with respect to the information delivered by capturing technologies (e.g. an vibration alarm could be sent to the user when he grasps a wrong tool). Although haptic feedback is often used to replicate real-world interaction forces, haptics has also the potential to provide cues that are not available in the physical world.

This kind of feedback can be presented to the user for example by applying vibrotactile stimuli to the human body or rather body part. Several studies have been conducted concerning amongst others the distribution of the stimulus actuators on the human body, decisions which kind of motors are better suited, and tests for detecting proper frequency and amplitude settings of feedback signals [SHPH06].

These studies have shown that cylindrical motors outclass pancake types, due to the vibration direction and their operating vibration frequency range. The maximum stimulus is perceived by the human skin at 250 Hz.

Liebermann et al. [LB07] developed a wearable vibrotactile feedback suit. They use this suit to provide real-time tactile information to a user while training motor skills. The suit analyzes the target movement, that can for instance be performed by the teacher, and applies real-time corrective vibrotactile feedback to the trainee's body in order to support him in acquiring the motor skills. Another example for a device applying vibration stimuli to the human skin is described by Schätzle et al. [SEWP10], who developed a vibrotactile bracelet that presents vibrations to the user arm, forearm, or wrist. There are many approaches for providing vibrotactile feedback to human body parts, such as the application of vibrotactile stimuli to the abdomen using a belt [BDC*10], to the shoulders by using shoulder pads [TDTA03], or to the torso using a vest [DCY10].

An exemplary device for guiding a human operator's movements by applying forces to a body part is presented by Bergamasco et al. [BAB*94]. The authors introduce an exoskeleton system that replicates external forces to the human operator's arm. This exoskeleton consists of a 7 DOF mechanical structure wrapping up the whole human arm and is directly supported by the shoulders and the trunk of the human operator. Experimental test have shown that the sensation of forces perceived by the human operator is acceptable and surely sufficient for the functional performance of primitive pushing and explorative tasks. More details about the use of exoskeleton systems as interfaces for teleoperation issues can be found in [BFA07].

4.4.3. Rendering of Audio Information

During training it can be useful to provide audio information, such as pre-recorded speech instructions or spatial sounds, for slightly guiding the user. That is, a spatial sound may be applied to show the user where he should look or move, or to direct him to the object of interest. The sound can be presented for instance via binaural headphones or a set of speakers.

Tsingos et a. [TGD04] propose a real-time spatial audio rendering pipeline for rendering audio samples. This pipeline consist of four steps which are repeated for each audio processing frame (i.e. every 20 to 30 milliseconds). In a first step, the perceptual saliency of all sources in the scene is evaluated. All sources are sorted with respect to their binaural loudness and subsequently perceptually inaudible sources are culled. The remaining sound sources are divided into a predefined number of clusters by applying a dynamic clustering algorithm which also considers the loudness of each source. Then, for each non-empty cluster a representative point source is created. In a next step, for each cluster an equivalent source signal is generated. Several operations on the original audio data, such as filtering, re-sampling, and mixing, are involved in this process. The source signals are used to feed the available audio hardware channels. In the last step, the pre-mixed signals for each cluster and their representative point source can be either rendered in software, or fed into standard audio rendering APIs (e.g. DirectSound3D, OpenAL).

4.5. Summary

In this chapter, an overview of technologies that are useful for development of Augmented Reality-based training systems has been provided. It becomes apparent that such training system should include both capturing and rendering technologies. Furthermore, according to the needs of the application domain and of the task scenarios appropriate display devices should be used.

Based on the information provided in this chapter and the preceding one, a training concept for multimodal Augmented Reality-based training can be developed. This will be done in the following chapters.

5. AR-Based Multimodal Training

For the development of a concept for efficient training of maintenance and assembly skills, it is important to analyze which technologies can be used to support the user during training and how they should be applied to support him best. Another crucial factor is how information and feedback are presented to the user during training. In this chapter, the use of multimodal feedback and Augmented Reality technologies for training is examined. Furthermore, a visualization concept for providing information in Augmented Reality training environments and a concept for the design of a multimodal Augmented Reality-based training platform are presented.

5.1. Multimodal Training

It has been demonstrated in Chapter 3.5 that receiving information with more than one modality can improve the human perception and enhance performance. Also the use of Augmented Reality technologies in the field of training has proven to be suitable in previous research (see Chapter 2). Due to these findings, the provision of multimodal feedback and Augmented Reality during training is analyzed in further detail.

5.1.1. Training with Multimodal Feedback

In previous research it has been pointed out that the user easily learns how to use feedback in different modalities and that humans enjoy using it. Studies showed that presenting multimodal information about the matter the person is studying can improve learning [May01]. For example, the performance of people who studied materials using audio text and visual diagrams or tables was better compared to people who studied with only visual information available [TFCS97].

According to Wickens' so-called Multiple Resource Theory, by using multiple information channels an optimal utilization of the user's cognitive resources can be achieved [Wic08]. A second model advocating the use of multimodal training is the Cognitive Load Theory, which argues that the use of several modalities reduces the cognitive load on the working memory [vMS05].

An important aspect that should be considered when developing multimodal training systems is the dominance of different modalities. Since vision naturally dominates over the other senses, it is often considered as the prime and preferred sensory modality for humans. For this reason it can dominate the experience. In simple and complex perceptual-motor tasks, for example, vision turned out to be the dominant modality [Bat54]. In simple motor tasks in which the visual search is focused only on the execution of a dedicated action, the visual dominance increases with the level of experience [PMGD87].

But it cannot generally be assumed that vision is always the dominant modality. Under certain circumstances other modalities can be predominant as well. An important factor for the dominance of modalities is the reliability of the stimuli inputs. How the user combines the perceived multimodal information depends on the level of reliability of the corresponding stimuli [BA06]. Also with proper training the level of dominance may be changed. For example, this becomes apparent when observing the acquisition of touch typing skills.

Also the presentation of audio information can be useful for training. Besides the content of audio information (e.g. error alerts, spoken instructions, a metallic sound on contact with a virtual metallic object), also the spatial component of presenting audio information may provide useful features for skill acquisition. So, for instance, the user can perceive simultaneously different spatially arranged sounds. Furthermore, since humans can detect audio signals faster than visual signals, a spatially arranged sound can very quickly direct the user's attention to a certain location or object, or can help him to orientate himself [EBW97]. However, spatial sounds cannot only be used to substitute but also to supplement visual information, which can improve and accelerate searching tasks [BDM99]. Also in situations in which the user is heavily loaded with visual information (e.g. a pilot controlling the aircraft in the cockpit), spatial sounds can improve the user's awareness of the situation and reduce his reaction time (e.g. [OVB04], [PSS*04]). Besides, the presentation of information in one modality (e.g. sound, tactile information) can immediately draw the user's attention to other sensory stimuli presented at the same location (e.g. visual information) [SPD04].

As already mentioned in Chapter 4.4.2, the haptic modality has the potential to provide information and hints that are not available in the physical world. In particular, haptics can be used as a channel for presenting motor patterns that a user is expected to internalize and later recall. Experiments demonstrated that humans can better memorize instructed force patterns presented both visually and haptically than patterns presented either only visually or only haptically [MTB*07]. Hence, the combination of visual and haptic modalities may lead to effective training. Furthermore, providing error feedback via the haptic modality (in particular vibrotactile stimuli) can provide the information directly at the location of interest (e.g. the user's left or right arm).

To summarize, the integration of different modalities in a training system can improve the training enormously. Since vision is in the majority of cases the primary modality, the availability of visual information and a qualified representation are crucial factors for the development of efficient training systems. As mentioned in Chapter 3.3, guiding and instruction are important for training, but must not impede the user's active exploration of a task. Hence, the different modalities should not be combined to amplify the guidance, but rather to create "channeling" information that supports the user in exploring the task and guides him slightly towards the goal. Based on the finding that the presentation of multimodal information improves performance, a multimodal augmentation of reality can and should be used to support trainees in performing the training tasks.

5.1.2. Training with Augmented Reality

As mentioned in the previous section, providing visual information during training is fundamental. By applying Augmented Reality technologies, the reality can be augmented by visual information. This information can be directly attached to real objects and accordingly the information can be linked to the objects of interest. In this way the representation of the visual information is enhanced by the spatial component of the information. The involvement of the spatial component in the visualization of the information supports the user in identifying the object of interest. By attaching visual information to an object, a direct relationship between information and target is established facilitating the interpretation of information for the user.

Besides, Augmented Reality can provide information sources and response capabilities that do not exist in the real environment of the task. For example, arrows can be displayed showing the user where to move his head to find the object of interest or 3D animations illustrating how to remove a machine part can be superimposed onto the real sensor (see Figure 5.1).

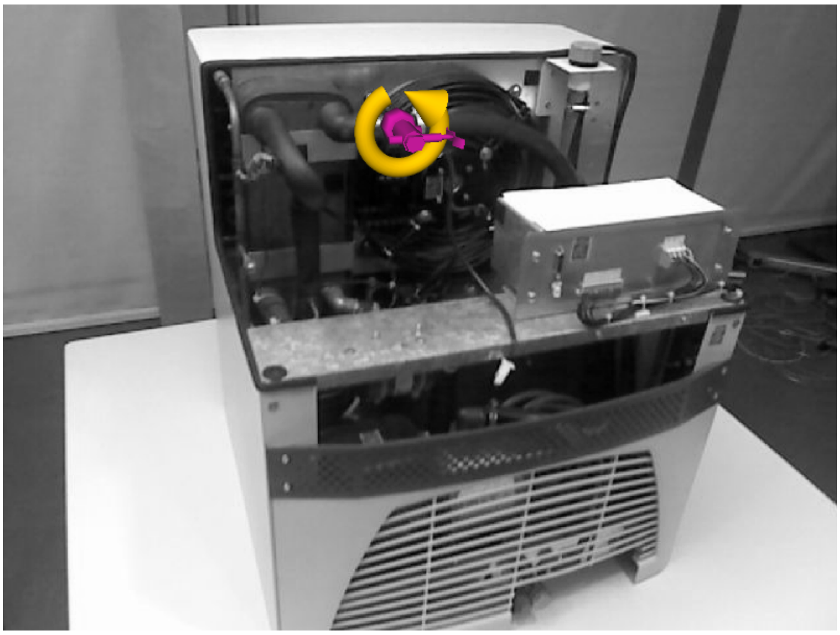


Figure 5.1.: A 3D animation shows how to remove a machine part.

One great advantage of using Augmented Reality for training is that the trainee can interact with the real world objects and simultaneously access the digital information for guidance (i.e. instructions, hints). Thus, the trainee can easily accomplish the mapping between the training and the real task. Furthermore, he can perform the actual task while accessing additional training material and that way learning is facilitated. With regard to Fitt's model of skill acquisition (described in Chapter 3.2.1) this denotes that Augmented Reality enables the trainee to learn the basics about the task by observing the augmented instructions and trying to perform the instructed sub-task, to develop behavior and movement patterns when performing the sub-tasks, and to redefine those motor patterns in repeated performances of the task (i.e. during training). By accessing augmentations while performing the task training, the trainee becomes partially skilled. His skill level starts to develop when he performs the real task for the first time.

Another advantage of Augmented Reality-based training is that the trainee has real tactile feedback when performing the training task, since he can interact with real objects. The virtual objects provide additional information about the task and its performance and supplement the trainee's knowledge about the task. Accordingly, the trainee can access the training material (i.e. the virtual instructions, etc.) and the real environment without the need to use "external" separate training material (e.g. a user manual).

Furthermore, the use of a training platform that involves virtual elements, as it does an Augmented Reality-based platform, allows for the measurement and evaluation of the trainees performance in modes and levels of detail that are not possible when performing the actual task in real world without using virtual components. By involving a technology providing virtual elements, it is also possible to respond to the trainee's performance and present corresponding feedback in a way that is not available without this technology. In addition, the type and order of presented sub-tasks can be adapted, which is not always possible in the real world.

As mentioned in Chapter 1.1, a potential danger of Augmented Reality applications is that users become dependent on Augmented Reality features such as visual (virtual) instructions. As a result, the user might not be able to perform the task when those features are not available or when the technology fails. This leads to two demands on Augmented Reality training applications and programs: On the one hand, the training programs should include phases in which the amount of AR features is reduced (less virtual components, e.g. only instructions for the current sub-task without additional information about the device, tools, etc.). On the other hand, the training program should also include phases in which the level of information provided by the AR features is reduced (e.g. only spatial hints without detailed instructions). That is, the level of guidance in the training system has to be adaptable to the current training phase. To summarize, AR-based training applications must clearly differ from AR-based guiding applications, as they must really train the user and not only guide him through the task. This can be only achieved by involving cognitive aspects in the training.

5.1.3. Potential of Vibrotactile Feedback

In daily life, haptic feedback is an important source of information. It is involved in almost every motor task performed by a human such as the grasping and manipulation of objects. To date, haptic feedback is integrated in a lot of devices and applications. Examples are mobile phones with vibration alarm and touch interfaces with tactile feedback [Uni06], force feedback joysticks used in computer games, training devices for laparoscopic surgery [Bau97], or human-machine interfaces for telerobotics [HSA*08].

As mentioned in Chapter 4.4.2, haptic feedback can be divided into two categories: kinesthetic and tactile feedback. Kinesthetic feedback transmits, for example, information about applied forces, torques, or guided movements, while tactile feedback corresponds to a mechanical interaction with the user's skin. Tactile displays present feedback for instance in the form of vibrations or pressure [SEWP10]. The so-called vibrotactile feedback belongs to the category of tactile feedback and is generated with devices that apply vibration stimuli to the human skin.

There are different research groups working on tactile feedback, corresponding devices, and the use of vibrotactile feedback in training (e.g [TH09], [WDF*10], [SEWP10]). Amongst others, it has been proposed to use vibrotactile feedback for communicating feedback about collisions [SEWP10]. Other studies focus on using tactile feedback to direct the user's attention. Attention direction can be useful in complex work contexts in which additional visual supportive information would lead to a too high load of the user's attention resources. Especially when the workload is high or multiple tasks have to be completed, vibrotactile stimuli provide better support [BPG*06]. It has also been pointed out that tactile information supplementary applied with visual information yield better task performances and shorter reaction times [BPG*06]. In particular the supplementary use of tactile feedback for attention direction has proven to be useful [ECR09]. It was also found that the presence of vibrotactile feedback has a positive effect on the user's situation attention [WDF*10]. Recent devices such as vibrotactile bracelets [SEWP10] offer also the possibility to present motor patterns and movement hints to the user, such as rotational or translational movements indications ([CEG*09], [WBE*11]).

However, the field of vibrotactile guidance can be basically distinguished into three different categories [WSH*11]:

- Attentional guidance: Vibrotactile stimuli can draw the user's attention to a certain event or to a particular body part (e.g. [WDF*10]).
- Movement guidance: Vibrotactile information can be used for guiding movements of body parts, such as arm movements. For example, this can be useful for correcting postures or enhancing training of motor skills (e.g. [vdLSB09]).
- Spatial guidance: The use of vibrotactile stimuli to guide users to a specific target has also turned out to be effective (e.g. [BDC*10], [EvERD10]).

It becomes apparent that the use of vibrotactile feedback is versatile and useful to enhance the level of intuitive perception of the environment and to support the user's interaction with the environment.

5.1.4. Development of Training Programs

The main objective of a training program is to guide trainees to make the best use of their abilities, maximize task performance, and increase training efficiency. Within a cooperation with human factors researchers and cognitive scientists the main aspects for developing training in multimodal environments have been elaborated based on literature review and experience in training. In this context we found that the success of a training platform and an accordant training program in achieving the described goals depends on five major aspects, which are mandatory components of a good training program (see also [GKG10]):

- The specification of the training task, the skills to be acquired, the objectives of training, and the criteria for graduation: This specification has to clarify exactly what should be learned, which goals should be achieved, and when they are achieved (i.e. how can be decided, if the desired goals are reached).
- The development of task scenarios, task versions and difficulty modifications that best represent typical emergences and key requirements of the task: This is based on the finding that an enhanced and variable training environment affords the development of a more flexible and higher level competence (e.g. [Gop07]).
- The definition of the key response and performance measures, and the determination of criteria for evaluating the trainee's progress concerning task performance and competence level.
- The development of designated feedback indices and knowledge of results information provided to the trainees, and the definition of their frequency and mode of presentation: This includes the implementation of different presentation modes of the information.
- The transfer of training, or rather the transfer from training to the actual task performance: This is determined by the relevance of the training experience in the learning environment. Apart from this, training systems are usually only part task trainers and do not completely represent the operational task which they coach. Hence, it is important to identify the relevance of this partial task and evaluate how it can be integrated in the entire operational task in real world. By using an Augmented Reality-based training environment, the relevance of the training experience is very high, since the training can be performed directly on-the-job in the real environment (i.e. the actual task can be performed during training). Hence, the transfer from training to actual task is facilitated. For the same reason, also the transfer and integration of the partial task into the operational real task is eased.

The definition of task scenarios and versions, the identification of key response and performance measures, and the development of feedback indices can be covered by developing *accelerators* and *training protocols*. In this context, *accelerators* can be understood as variables that are inserted in the training process in order to facilitate, enhance and improve learning [GKG10]. They involve different modalities, such as vision, audio, and haptics. A detailed description about the concept of accelerators will be given in the next section (see Chapter 5.2). The schedule and duration of training, the selected task scenarios, the different difficulty stages, and the order in which they are applied are defined in the *training protocols*. An evaluation of accelerators and training protocols is required to assure the best use of the developed platforms as training devices. Accelerators should be evaluated in their contribution to performance and learning. Training protocols should be assessed to determine the number, duration, and structure of training sessions, as well as content, composition, and order of exercises and task scenarios (see also [GKG10]).

Furthermore, the development and design of task scenarios and versions (point 2) and of feedback indices (point 4) form the building blocks for learning, transfer, and for the design of the accelerators. The development and systematic application of these elements are crucial factors for developing efficient training programs.

For the evaluation of the developed skill training platform and for the assessment of how to apply it best in training, three basic evaluation aspects need to be analyzed: (1) the evaluation of the value of the accelerators; (2) the assessment of the training protocols in order to optimize learning and skill acquisition on a platform; (3) the transfer of the training studies to other similar tasks (see also [GKG10]).

5.2. Accelerators

As mentioned, the term *accelerator* describes variables that are introduced in the training process to improve and accelerate training. We introduce this concept of accelerators in order to define and evaluate how technologies can be applied to improve training. The main objective of accelerators is to guide the user to make the best use of his abilities, improve task performance, and increase training efficiency. In the context of multimodal Augmented Reality-based training, this implies that an accelerator defines how one or more of the available technologies can be used in order to (1) reduce the learning/training time and/or (2) improve the the trainee's skill performance.

Generally, we divide the accelerators into three categories ([SKI11]): augmentation, simplification and variability accelerators (see Figure 5.2).

5.2.1. Augmentation Accelerators

Augmentation accelerators are variables that enhance the training by providing additional information about the task and the task performance, and hence they augment the training. We further distinguish augmentation accelerators into four sub-categories:

Performance feedback accelerators provide information either about the trainee's performance regarding the task or an addressed sub-skill, or about the performance history. The information is mostly provided visually. For example, such an accelerator can provide feedback about how well the trainee is currently performing the task or how well he has performed overall so far (history feedback). Figure 5.3 shows a performance feedback accelerator. Two scores are visualized numerically and graphically. One is indicating the score of the current performance, and the other one shows the score of the overall performance. The colors of the progress bars correspond to the quality of performance; green signifies a good performance, and red signifies a bad performance. Different parameters can be mapped to these two scores. For example, for the training of singularity avoidance in programming a robot the first score of this performance indicator represents the current singularity index of the robot (i.e. how close the use is to a singularity), and the second one shows the total score of the trainee (i.e. how well did he

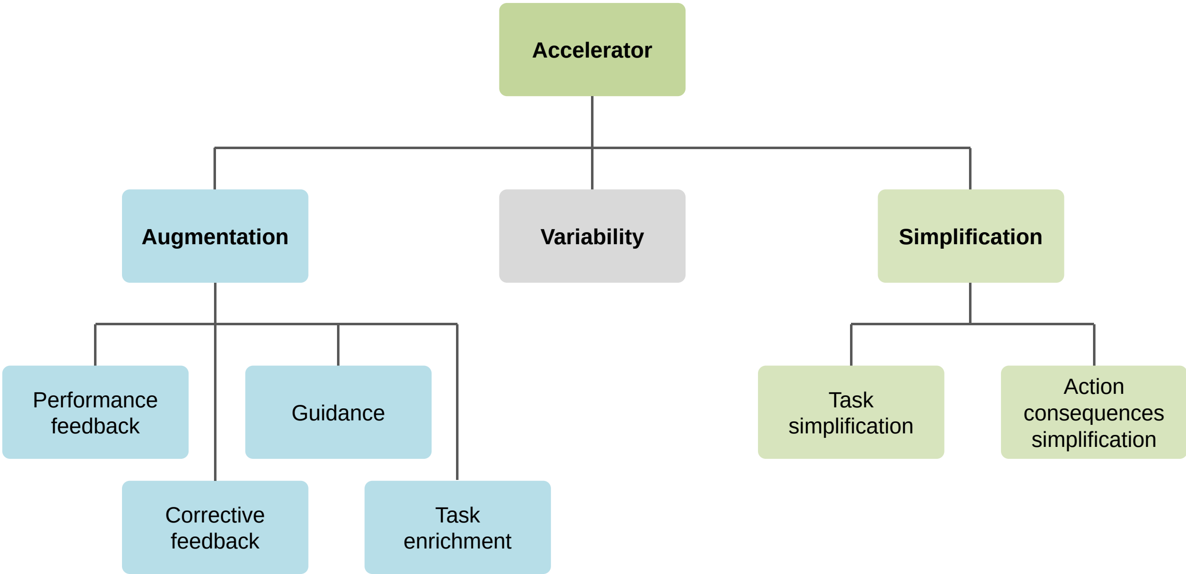
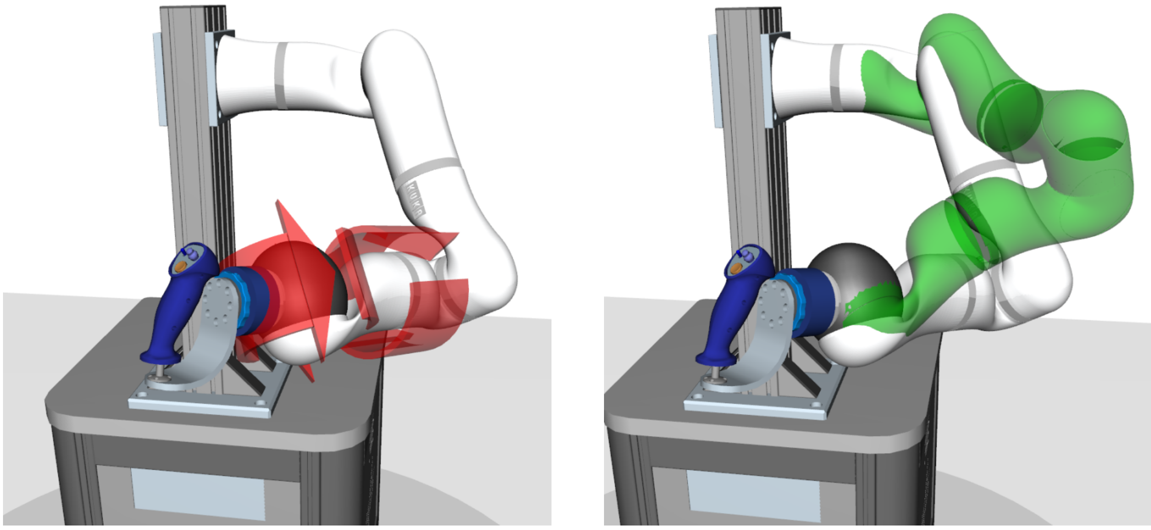


Figure 5.2.: Overview of the different accelerators.

avoid singularities).

Figure 5.3.: Performance feedback accelerator indicating the current and overall performance of the trainee. The colors correspond to the score.

Corrective feedback accelerators are elements which present feedback to the user with the aim to correct his performance. The feedback consists of visual, audio, or haptic information (or a combination of them). For example, haptic feedback can be provided to prevent the trainee from entering error regions or executing wrong actions. Also vibrotactile or audio feedback can advise the trainee of performing falsely (e.g. a vibration at the trainees arm and a failure sound can communicate that he grasped the wrong tool). Another example is shown in Figure 5.4 (a). A rotating arrow shows the trainee how he should move the robot joint in order to avoid a singularity. Corrective feedback accelerators can additionally provide information about the magnitude of difference to an optimal task performance. For instance, an animated transparent "ghost geometry" of a robot demonstrates the user how to move to avoid a singularity (or get out of it) and also how far the robots is away from an appropriate constellation at this moment. That way, also the difference between the current constellation and a better constellation is presented (see Figure 5.4 (b)).



(a) (b)

Figure 5.4.: Corrective feedback accelerator: (a) a rotating arrow indicates how to move the robot joint to avoid a singularity; (b) an animated transparent "ghost geometry" indicates how to move the robot arm to avoid a singularity and which constellation of the robot would be better in the current situation (rendered with *instantReality* [Fra08], screenshots made by DLR).

Guidance accelerators provide instructional information to guide the trainee through the task to the designated goal, which is an important aspect for training as described in Chapter 3.3.2. They can give directions on what to do by presenting the audio, visual, and haptic aids. For instance, visual instructions can be displayed to show the user how to perform a task or haptic guidance can lead the trainee to a dedicated direction. Guidance accelerators can also be attention highlighters, such as a spatially arranged sound bringing the user to move his eyes to a certain target or a pulsing visual shape highlighting an object (see Figure 5.5). Besides, they can display the consequences of actions, such as a visualization of future trajectory.

Task enrichment accelerators visualize information that could not be seen otherwise in order to improve the user's comprehension of the task. Examples are virtual replications of occluded real objects overlaid at the proper position, or a superimposed transparent 3D model showing a device after a successfully performed training task (see Figure 5.6).

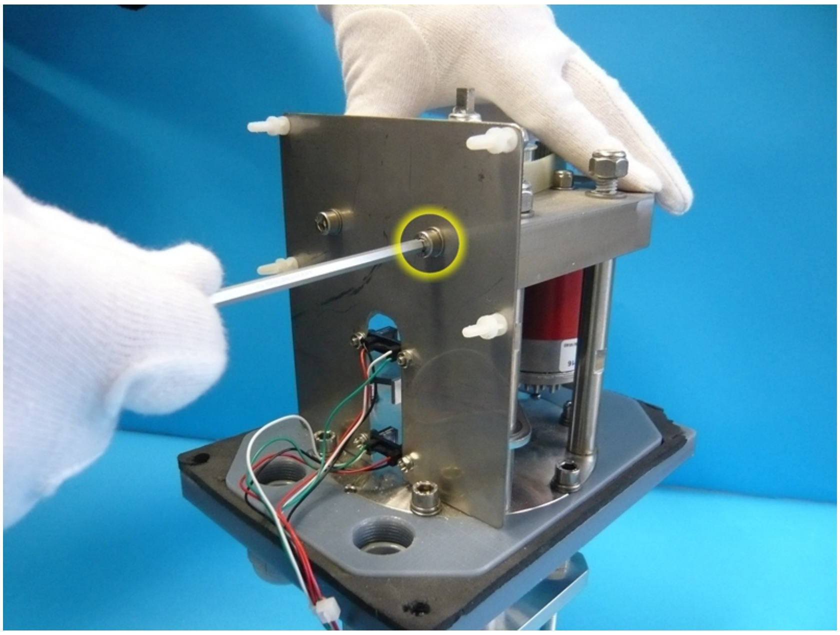


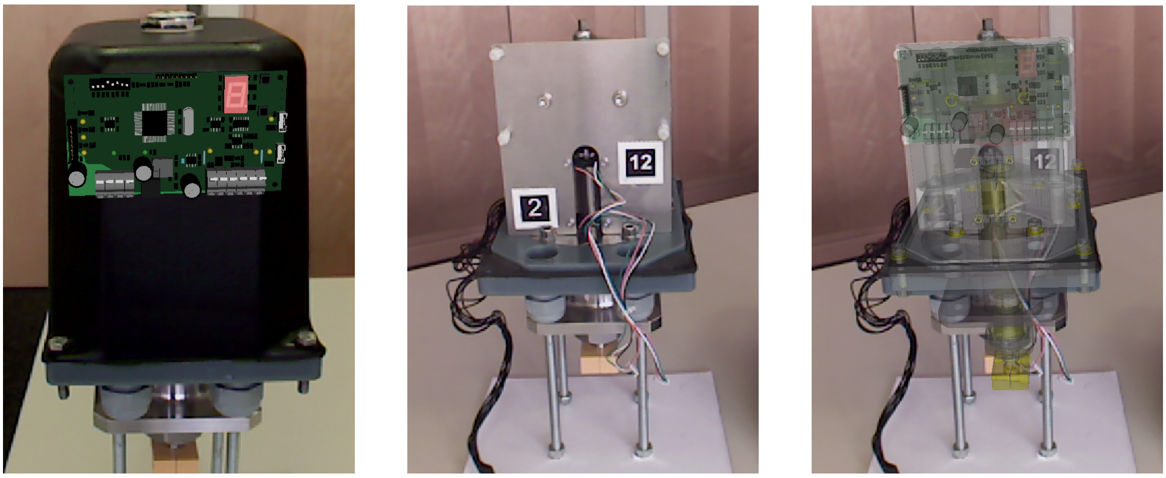
Figure 5.5.: Guidance accelerator: a pulsing circle highlights the object of interest (i.e. the screw).

5.2.2. Simplification Accelerators

The second category are the *simplification accelerators*. Those accelerators are simplifications of the task, of the consequences of actions, or of the environment. Examples for the first two groups are a kinematic simplification of a robot, error compensation (e.g. tremor reduction, motion amplification), and task simplification (e.g. haptic snap, training of parts of the task, emphasis of parts during task training).

The simplification of the environment (e.g. gravity reduction, slowing down of the simulation time) refers to Virtual Reality environments. For instance, if a person is trained in juggling using a Virtual Reality training system, the simulation time can be slowed down in order to give the trainee more time to coordinate his movements and interact with the balls.

As mentioned in Chapter 3.3, the simplification of difficult training task in early training stages facilitates the practice of the task for the user and helps him acquiring competence. Hence, introducing simplification accelerators in early training phases can improve training.



(a) (b) (c)

Figure 5.6.: Task enrichment accelerator: (a) an overlaid 3D model of the occluded circuit board; (b) the actuator before attaching the circuit board and without a virtual augmentation; (c) a transparent 3D model showing the device after a successfully performed assembly task, i.e. after attaching the circuit board.

5.2.3. Variability Accelerators

The third category, *variability accelerators*, introduce different training conditions into the the training program, such as the generation of different error conditions, varying visualizations of instructions, or different time modes of presentation of information (continuous, intermittent, or on demand).

It has been described in Chapter 3.3 that introducing variability in training is an important factor for supporting the trainee in acquiring robust and flexible skills. Since complex tasks usually are variable in difficulty, workload, and demand, the trainee needs to learn to adjust his response to cope with such changes (see [Gop07]). The integration of variability accelerators in the training can enhance learning and improve the trainee's ability to generalize the task and to transfer it to changed conditions.

As already mentioned, accelerators should be evaluated in their contribution to performance and learning. For the evaluation of accelerators, three important aspects should be considered: (1) the relevance of the information (content and type) to performance and learning; (2) the ability to improve learning without developing dependence on the information availability; (3) the question, if the provided information distracts the trainee or interferes with the natural modes of performing the trained task. The first aspect is only important for information providing accelerators.

5.3. Visualization of Information

Based on the finding that humans (who are capable to see) absorb most of the information through the eyes and that vision dominates the modalities in the majority of cases, the visualization of information is a crucial factor for the development of training systems. In this chapter an approach displaying detailed information (e.g. text) in interactive Mixed Reality environments is presented. Furthermore, a new concept for visualizing Augmented Reality overlays is introduced, which allows for controlling the presented information level of the overlay and for the integration of different types of media.

5.3.1. View-Aligned Visualization

In order to support the trainee during training and to avoid confusing him, the visual information presented to the trainee during training must be clearly recognizable and easily locatable. Particularly when context information, structural information, or any other type of information without a direct spatial relation is presented, the user needs to have direct access to the information without having to "search" for it (or to navigate to the information), which would have a negative impact on the performance time. Also if the content of the presented information is not clearly recognizable for the trainee, the performance time is impaired (e.g. displayed information is jittering, text is to small because far away from the scene camera).

A good way to overcome these problems is to align the virtual objects that visualize the information with the user's view. That is, they are rendered in such a way that they always appear parallel to the viewing plane in the user's field of view. This can be realized by placing the virtual object into the coordinate system of the scene camera (see Figure 5.7). When the view changes (e.g. HMD: the user moves his head; hand-held: the user moves the device with the camera; VR: the user navigates through the scene) the virtual objects still remain at their position on the screen. An example for a view-aligned visualized information is shown in Figure 5.8 (a). A progress-bar at the top of the screen provides information about the user's progress in the entire training task. If he moves the camera, the progress-bar remains at its position.

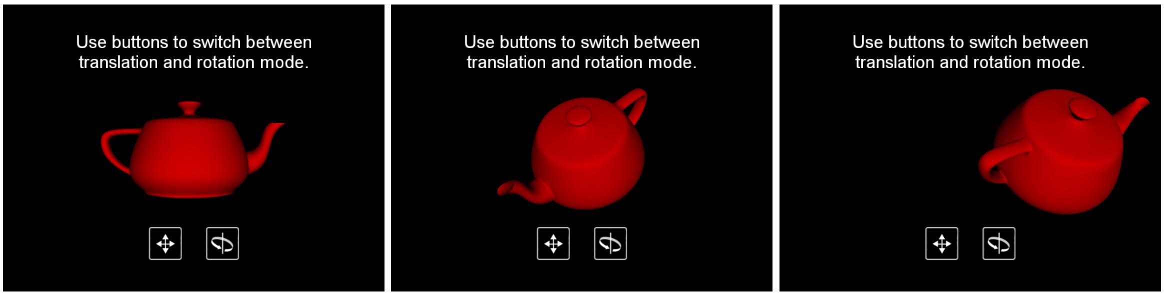
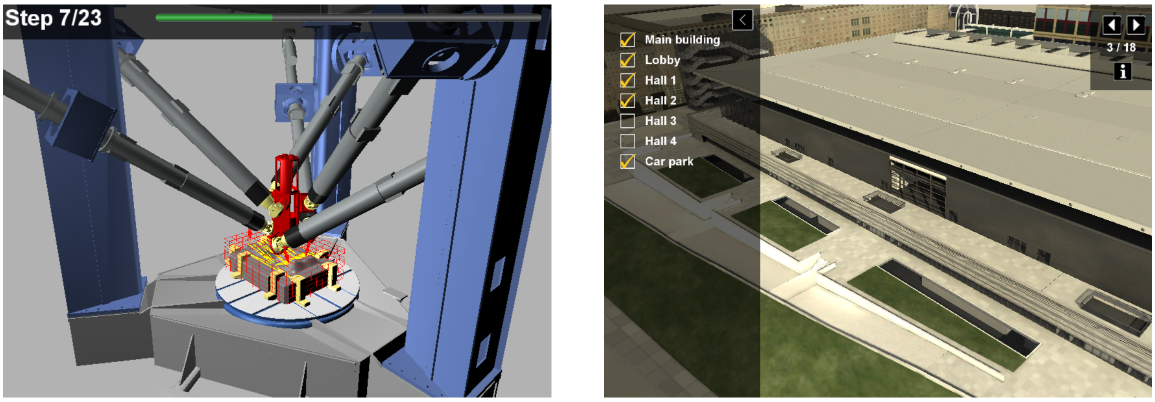


Figure 5.7.: View-aligned objects (text, buttons) are placed in the coordinate system of the camera. They are not affected by the modification of the scene camera.

In particular for Augmented Reality-based training applications, the view-aligned visualization of certain objects can bring enormous benefit. In real world, a human moves his head almost permanently. If he is told to stop moving the head and hold it still while performing any task, he feels highly uncomfortable and his movements become unnatural. In case the trainee is restricted in his natural behavior during training, the transfer of the trained task to the real world task is more difficult. To facilitate this transfer, the training conditions need be as close as possible to the real world conditions (cf. Chapters 3 and 5.1.4). That is, the user should not be restricted in his natural behavior. If a HMD is used as display device in an Augmented Reality application, the camera image changes according to the user's head, and as a result each overlaid object "moves" in the user's view. In case a hand-held display device is used, the trainee usually utilizes the AR device just to understand what to do, and then he puts it aside in order to perform the task with both hands. Hence, he is not restricted in moving naturally. However, if the device is comparatively heavy (e.g. a tablet PC), after using it a while the user's hands may start shaking when holding the device, and accordingly the camera image and the overlays start jittering as well. For these reasons information without a direct spatial relation (e.g. context information, structural information) and virtual objects containing very detailed elements (e.g. text) should be visualized view-aligned.

It is advisable to display view-aligned objects docked to the screen in order not to disrupt the view of the real world (and corresponding overlays) on the AR display. Besides, in many cases the information provided by view-aligned objects is not the main information, but rather a supportive or enhancing element, and hence it should not dominate the whole presentation on the screen. The level of conspicuity, and hence also the level of relevance, can be further controlled by adding transparency to the object. This can be further useful to avoid the total occlusion of a big area on the display, in case multiple view-aligned objects are displayed at the same time (see Figure 5.8 (b)).



(a) (b)

Figure 5.8.: View-aligned objects docked to the screen: (a) a view-aligned progress-bar; (b) transparent view-aligned objects (model controller, control panel); the main scene shines through the objects and is not totally occluded.

5.3.2. Adaptive Visual Aids

In Chapter 5.1.2 it has been pointed out that the training program should include phases in which the level of guidance is reduced in order to avoid the development of a dependency to this guidance tool by the user. In traditional Augmented Reality training systems the instructions are given either using overlaid 3D models or animations showing how to perform a task, or using text boxes attached to an object describing in detail how to manipulate this object. The information is always provided at a fixed, strong level of guidance.

Here we introduce the concept of *Adaptive Visual Aids* (AVAs). For this purpose the real-world concept of attaching annotations is adapted. A person annotating an article in a magazine or book usually marks or highlights the corresponding text and writes the annotations at the margin (if he annotates more than one word). Hence, the annotation consists of two parts: annotation *pointer* and annotation *content*. The annotation pointer can contain a variable amount of information depending on how much text is marked (a word, a sentence, etc.). Also the annotation content can comprise a different amount of information, depending on how much the person annotates. Within the context of this work, the described pointer-content metaphor has been adopted, transferred to virtual objects [Web08b], and further adapted for Augmented Reality [WBE*11] (see Figure 5.9).

Adaptive Visual Aids are composed of a spatially arranged pointer component and an optional content element providing additional detailed information (see also Figure 5.11). Both can provide a variable amount of information. The pointer consists of at least one virtual object overlaid on the camera image (like traditional Augmented Reality overlays). It presents also the spatial component of the information. The pointer object can consist for example of a complete 3D animation, a 3D model, or a highlighting geometry (e.g. a pulsing circle). Thus, it can directly visualize instructions (e.g. in the form of 3D animations) or just act as an area highlight that is marking the area of interest. The guidance level of the pointer is a continuum between a very strong guidance (e.g. a 3D animation) and a very soft guidance (area highlight). This is illustrated in Figure 5.10.

The content object is visualized view-aligned (tracking-independent) and docked to one side of the screen. It can be shown or hidden on user demand. It consists of a view-aligned virtual 2D plane and data objects in various media formats which are visualized on that plane. Thus, it can provide multi-medial information, that is clearly recognizable for the user. The data objects displayed on the plane can contain text, images, videos, and 3D scenes rendered in a 2D image on the plane, or any combination of those elements. That is, the content object can contain detailed instructions (e.g. a text description and a video showing an expert performing the task) or just a hint (e.g. a picture of the tool needed to perform the task). The amount of information visualized through the content object is also a continuum between no information (no content, no guidance) and a detailed description of the task (strong guidance), as illustrated in Figure 5.10. Examples for Adaptive Visual Aids with different guidance levels are shown in Figure 5.11.

As mentioned above, the visibility state of the AVA content object can be changed on user demand during runtime. If the information on the content plane is no longer needed or the user wants to use the

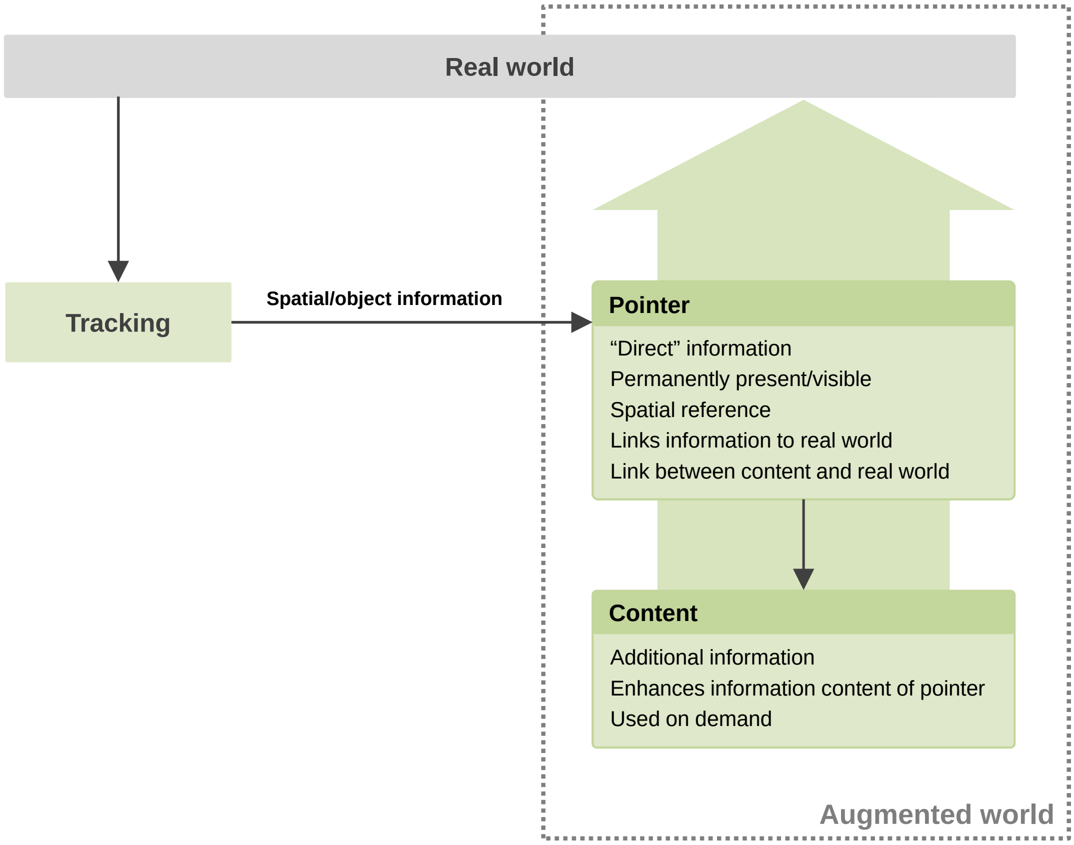


Figure 5.9.: Transfer of the pointer-content metaphor into Augmented Reality environments: due to the spatial and/or object information presented by the pointer, the information content (of pointer and content object) is linked to the real world.

whole screen to explore the training task, he can request for hiding the content plane.

Beside the support of different guidance levels, the use of AVAs has also further advantages. One is the possibility to present different types of media in the content component. Showing 3D animations is not always the best way to explain a task. Different tasks require different types of media content to provide the best possible support for the trainee. For example, if the trainee has to perform a task in which the manner of interacting with the device is crucial (e.g. the user has to remove a plug from a damageable board while fixing the board carefully with one hand in order to avoid to damage it), it may be more effective to show the user a short video that demonstrates exactly how to interact, than showing a complicated 3D animation. Furthermore, by using AVAs different media elements can be combined, which can further improve the information value. This theory is supported by the

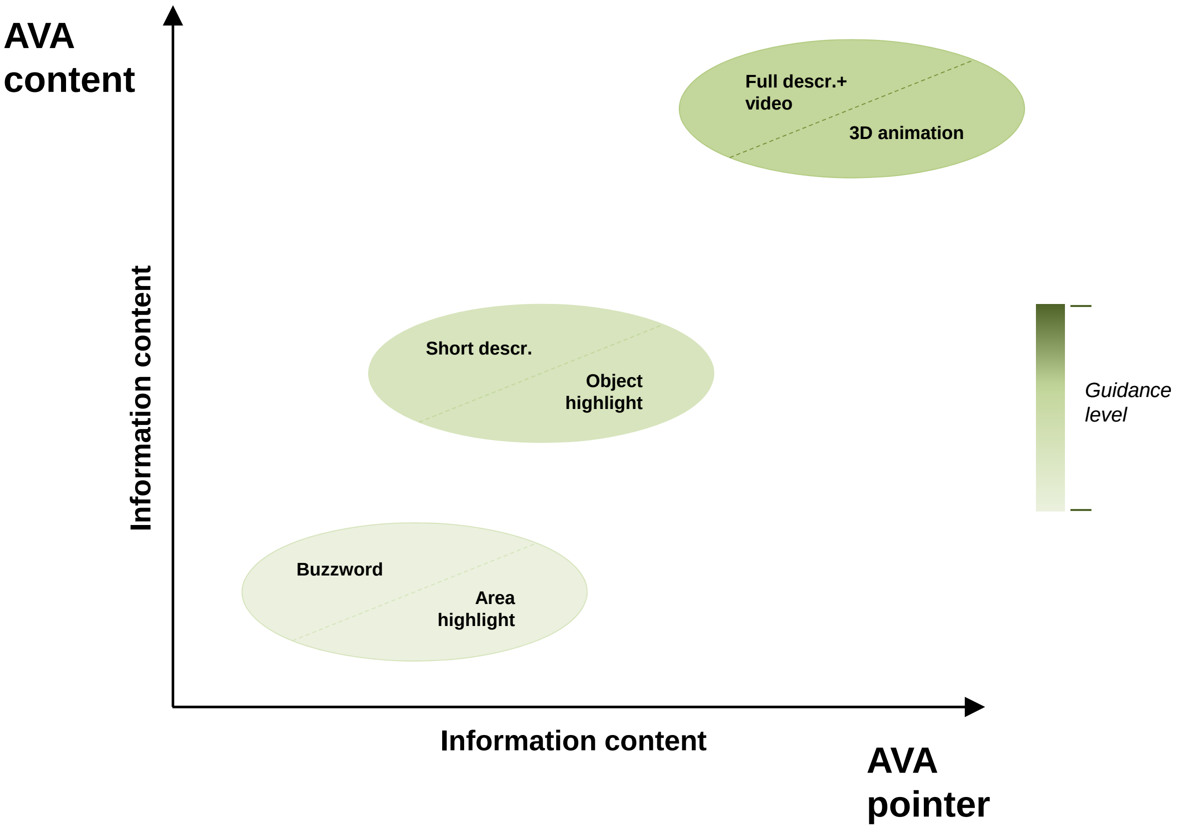
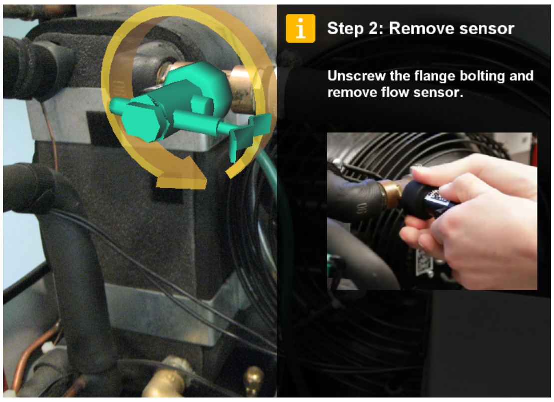


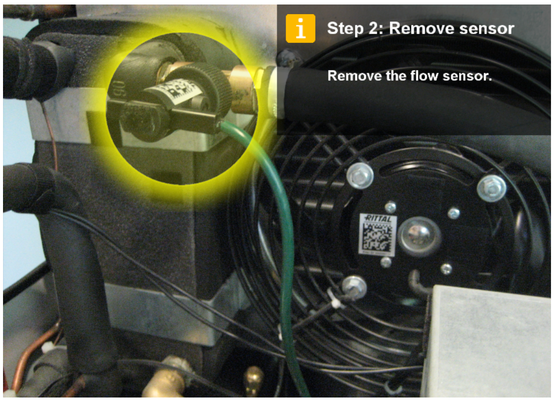
Figure 5.10.: Information continuum of Adaptive Visual Aids: the guidance level depends on the type and amount of visualized information and is a continuum between very strong and very soft guidance.

Multimedia Representation Principle, which states that "it is better to present an explanation in words and pictures than solely in words" [May02].

Considering the aspect of content generation, it becomes apparent that the process of content generation has always been a critical factor in the development of Augmented Reality-based training systems. In many cases the generation of the 3D instructions, such as 3D models and animations, is not a completely automatic process. Even if the 3D data is already available in a database and can be exported in the target format, it usually needs to be prepared before it can be used in the training system (e.g. polygon reduction if the system hardware does not provide enough processing power), which may cause a big effort. Shooting a video clip or taking a picture that demonstrates the task to perform or shows the object of interest causes much less effort. Usually a lot of training material already exists, as it is utilized in traditional training programs. Particularly if the training material has proven very good and useful in the traditional training, it should be integrated in the AR training as well. Since Adaptive



(a)



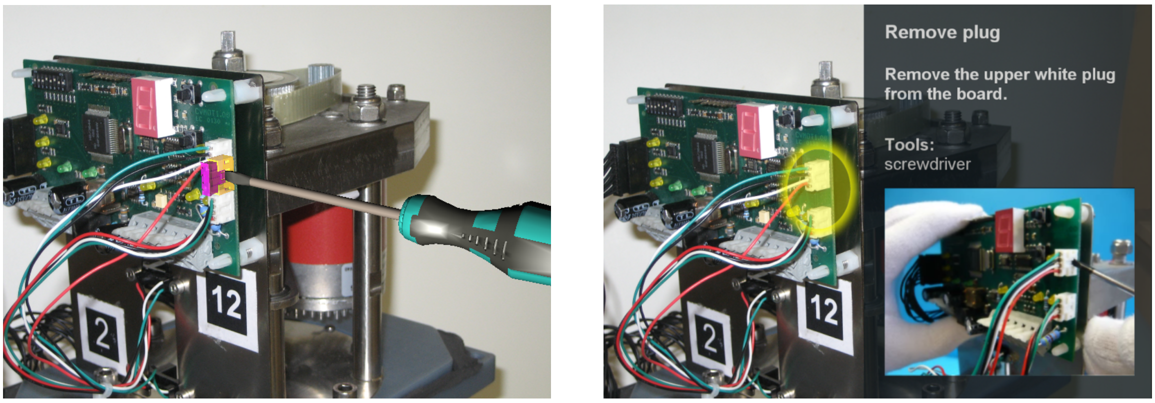
(b)

Figure 5.11.: Adaptive Visual Aids with activated content object (a) at a strong guidance level and (b) at a much softer guidance level.

Visual Aids allow for integrating multi-medial information, all this data can be used in the AR training system. The data can be stored in the visual aid object and visualized in arbitrary combinations in the content component of the Adaptive Visual Aid.

In particular if the training takes place in the target environment (and not in a training lab), it is not always allowed and possible to modify the environment and markerless tracking must be applied. As mentioned in Chapter 4.1, markerless tracking under real world conditions is still a challenging field, so a perfect registration in unprepared environments cannot be assured, and hence cannot be assumed by the training application. Several studies showed that registration errors harm the user's understanding of the overlaid information and have a negative impact on his task performance ([LA08], [RMW09]). Traditional training systems which use only overlaid 3D models and animations as instructions suffer a lot from this effect. In contrast, the use of Adaptive Visual Aids overcomes this problem: If a very accurate registration is not available, the pointer can be used to highlight the area of interest (e.g. pulsing circle around the area of interest), while the content element presents the instruction to the user and clarifies definitely which action to perform (see Figure 5.12).

Considering the described advantages, it becomes obvious that Adaptive Visual Aids are not only useful for training, but may also enhance traditional Augmented Reality applications, since information can be presented in the most suitable way while keeping the spatial relationship to the real environment. Furthermore, AVAs are a good possibility for presenting visual augmentations on small screens (e.g. ultra-mobile PCs, smartphones). A complicated 3D animation can be difficult to recognize on a small screen. In contrast, an Adaptive Visual Aid can additionally provide a bigger presentation of the information on the content plane using different types of media, such as a zoomed in visualization of the 3D animation or a video. Since the AVA content plane can be shown and hidden on demand, the user can decide when the additional information is necessary. If the information on the content plane is no



(a) (b)

Figure 5.12.: (a) The traditional AR overlay suffers from registration error; (b) the Adaptive Visual Aid can compensate the same registration error, since the AVA content object clarifies definitely which object to manipulate.

longer needed, he can hide it and use the whole screen for exploring the Augmented Reality application.

To summarize, the concept of Adaptive Visual Aids allows for presenting instructions at different information and guidance levels. AVAs can contain different kinds of media, such as text, images, videos, and 3D content, which can be arbitrarily combined. Thus, existing training material can be integrated in the training system. Also the process of content generation can be facilitated and accelerated, since it is not absolutely necessary to present instructions by using 3D models and animations. In addition, Adaptive Visual Aids overcome problems of traditional AR systems in case that a very accurate registration is not available.

5.4. Multimodal Training Platform

In the following, an architecture concept for a multimodal AR-based training platform is presented. This concept is not specific for the training of maintenance and assembly skills, but is a general design approach for multimodal training platforms. Hence, it can be adapted for training in different application domains such as rehabilitation and surgery.

5.4.1. Architecture Overview

Basically, an Augmented Reality-based multimodal training system exists of three main components: the physical training environment, the multimodal training platform, and a digital data repository (see Figure 5.13). The architecture does not only consider the training of users, but also the idea of being trained by experts. This is an important aspect, because by capturing the experts' behavior existing workflow descriptions of the training task can be refined (or newly created) in order to get the best description of the skill, or rather of the workflow that underlies the skill. For this reason, the architecture of the system includes modules for capturing and processing expert data.

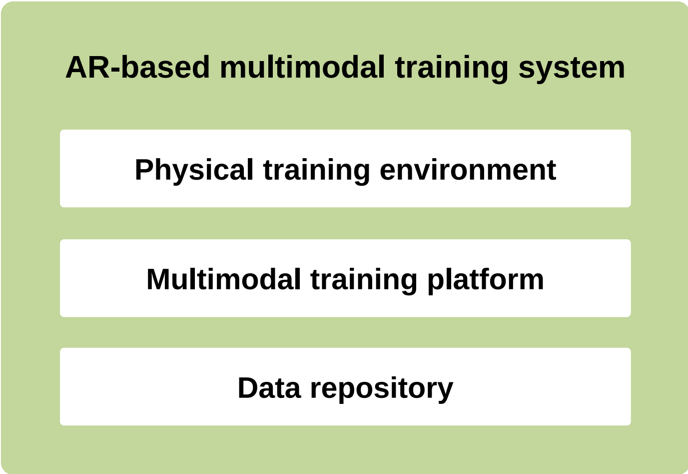


Figure 5.13.: The main components of an AR-based multimodal training system.

The physical training environment includes all physical objects involved in the training task as well as the real environmental conditions like temperature and lighting. Additionally, it contains the interfaces involved in connecting the real world with the computer such as for instance display devices, cameras, and haptic devices. For example, in a maintenance task training the physical training environment includes amongst others the physical machine to maintain, the necessary tools, new machine parts to build in, and the utilized display device.

The data repository is a kind of database storing data like workflow descriptions, training protocols, training material (e.g. visual aid objects, speech instructions, 3D models), templates for presenting feedback (e.g. alert sounds), templates for presenting vibrotactile hints like rotations or translations, etc.), data captured from experts, and user specific data (e.g. current training phase).

The multimodal training platform handles the capturing, the processing of the captured data, the selection of accordant instructions, the generation of adequate feedback, and the presentation of information and feedback to the user. It can access the data repository and has interfaces to connect to the physical components (e.g. cameras or display devices). The platform includes the whole logic of the training application, such as the determination of the next step and the selection of corresponding instructions and feedback regarding the current training phase. It basically consists of three major components: a multimodal capturing controller, an interaction processing and application module, and a multimodal rendering controller. Figure 5.14 shows an overview of the platform and its components. In the following the main tasks of these components will be described.

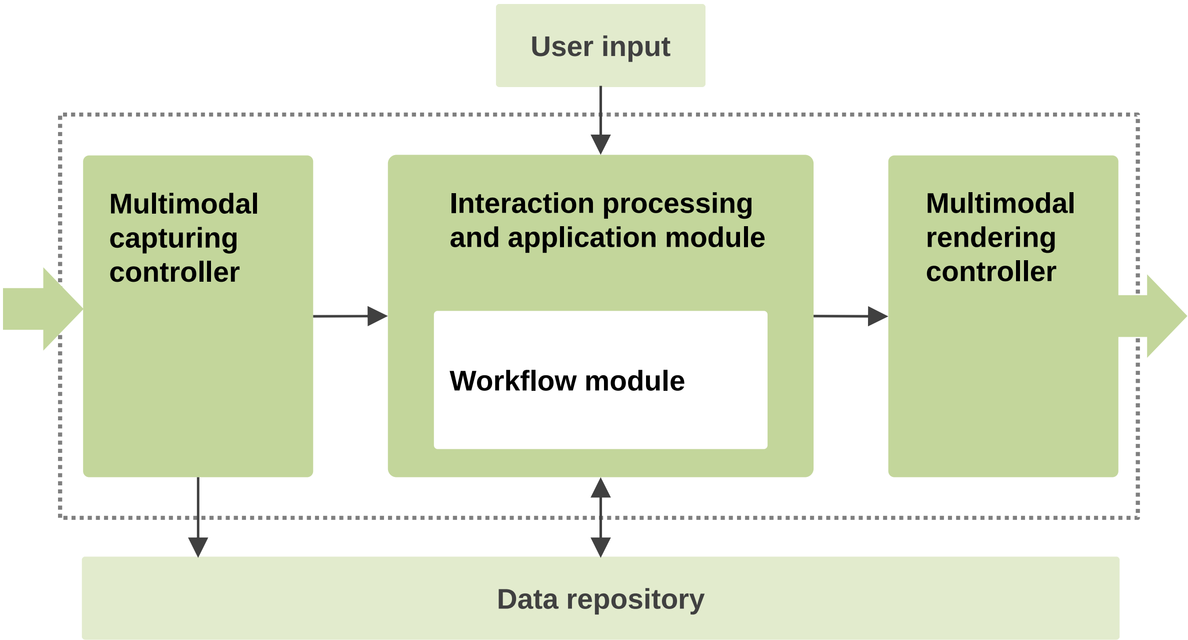


Figure 5.14.: Architecture overview of the multimodal training platform.

5.4.2. Multimodal Capturing Controller

The main task of the multimodal capturing controller is to efficiently handle the data delivered by the different capturing systems. This can include camera tracking for Augmented Reality, tracking of objects and hands, motion capture, capturing of biometrics (e.g. eye-tracking), and audio capturing, as well as EMG or capturing information about the oxygen consumption. The architecture of the capturing controller is shown in Figure 5.15.

The controller provides access to all necessary capturing devices such as cameras or inertial sensors. In addition, it handles all preprocessing of the data which is necessary to provide a simple and clear interface for accessing the data in the interaction processing and application module. Therefore, the data must be abstracted and a consistent interface must be defined to enable an uniform access to all kind of capturing data. Furthermore, all captured data must be synchronized and parallelized communication channels between the software components must be provided.

The capturing controller can also access the data repository. This is mainly important for storing the data captured from experts in order to be able to reuse it in different training programs. Also several tracking techniques exist that reuse existing tracking information (e.g. feature maps of the tracked environment as described in [WWK11]) to improve the performance and accuracy. This data is stored in the data repository as well. Moreover, the capturing controller allows for accessing data on different machines on the network. This is important, since multimodal training systems potentially

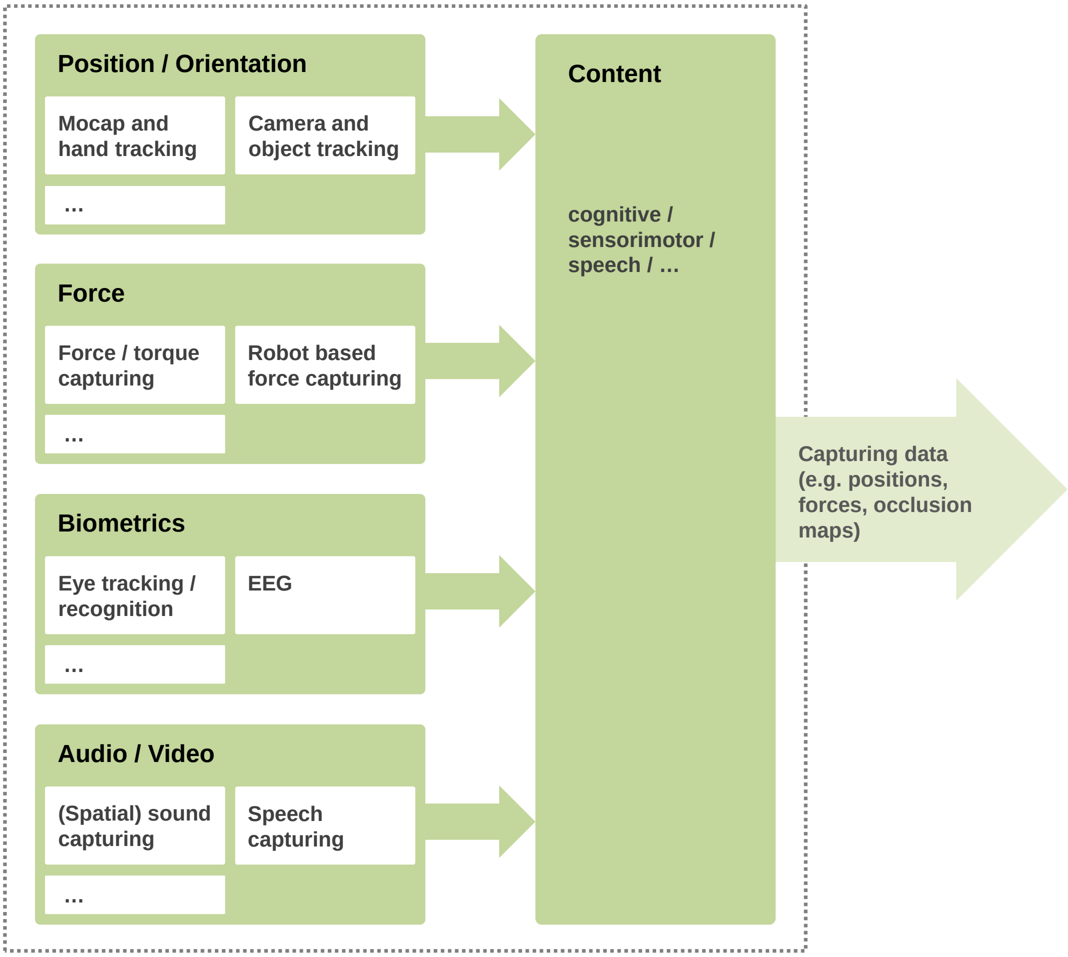


Figure 5.15.: Archtitecture overview of the multimodal training platform.

need to handle devices that are not connected to the machine the application is running on (stationary tracking systems that determine the position and orientation of a mobile device, stationary motion capture systems, etc.).

The minimum requirement that the capturing module has to fulfill is to send in each frame a tracked camera pose and the corresponding video image to the interaction processing and application module for generating the Augmented Reality overlays. This pose and all other available capturing data is also reported to this module for further processing. If occlusion maps, or rather depth maps, are available, they are passed through the interaction processing and application module to the rendering controller in order handle occlusions (e.g. real hand occludes virtual object).

5.4.3. Interaction Processing and Application Module

The interaction processing and application (IPA) module contains the application and workflow logic and processes the user input. It is connected with all components of the platform. In addition it gets input initiated by the user (e.g. request for more information via mouse click). An overview of the module is given in Figure 5.16. The information about the current step to process is given by the workflow block, which is part if the IPA module. The data delivered by the capturing controller is used together with data from the data repository (e.g. visual aid objects, 3D models, current training phase) to build the Augmented Reality overlays with respect to the current training phase and send the resulting visual data to the rendering controller.

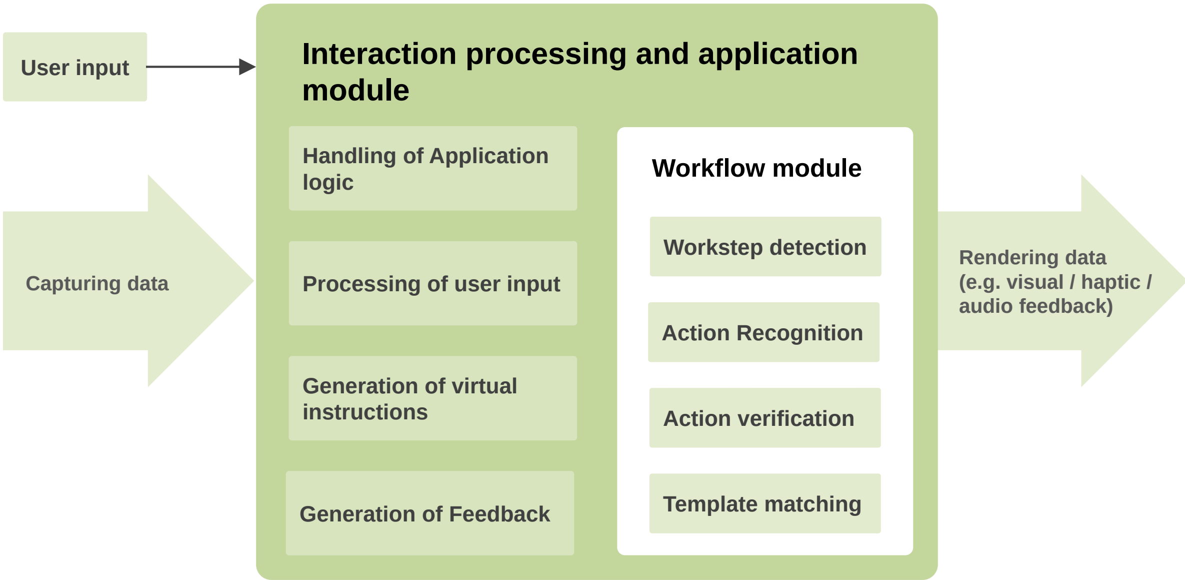


Figure 5.16.: Overview of the interaction processing and application module.

5.4.3.1. Workflow processing

As mentioned before, the module contains a workflow block which handles the whole workflow of the training task. It uses data stored in the data repository, such as workflow descriptions, to control the workflow of the training task and to check the user's performance based on input data from the capturing controller and input actively initiated by the user (e.g. triggering of next step via mouse click).

The workflow description defines how the single steps in the training task have to be passed and which requirements must be fulfilled (e.g. screw number 5 must be replaced using screwdriver 1). The simplest workflow description is a linear description that consists of one fixed sequence of steps to perform. The next step can either be triggered by the user or result from captured information. For instance, the user is told to press a button when he finished a step in order to get the next instruction. Based on this user input the next step is selected. This is the minimum demand the workflow block has to meet.

Alternatively to this simple workstep selection, the workflow block can recognize the action performed by the user and use this information further to decide the next step. For example, when the module recognizes that the user removed the screw, it triggers the next step. This is done by the *Action Recognition component*. This component uses the workflow description and input data from the capturing module (e.g. object recognition, hand tracking) to determine which action is performed. The workflow description delivers step descriptions or templates that result from captured experts or other competent users performing the current step, which serve as reference data for the recognition of the user's action. Furthermore, the magnitude of difference to the reference data can be calculated in order to generate corrective feedback for the user. There are different approaches for recognizing actions, mostly based on machine learning techniques, such as Dynamic Time Warping (e.g. [AAYS05], [AV10]) or a Hidden Markov models [WSB09].

To go in detail, a Hidden Markov model is a doubly stochastic process based on Markov chains [Rab89]. When an HMM is generating a sequence, a certain state is visited and an observation is emitted depending in the emission probability distribution of the current state. Then a new state is chosen depending on the state transition probability distribution. Thus, the model generates two strings of information: An underlying state path and a sequence of observations. Since only the observation sequence is observable, the underlying state path is hidden, i.e. the states can not be observed. In other words, the HMM extends the Markov Model by emission probability distributions. An HMM is characterized by the following parameters:

- The number of states N, denoted as $S = \{S_1, S_2, ..., S_N\}$
- The state transition probability matrix $A = \{a_{ij}\}$,

  $$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \quad 1 \leq i, j \leq N$$

  where q_t denotes the actual state at time t and a_{ij} is the transition probability of taking the transition from state S_i to state S_j.
- The emission probability matrix, that is $B = \{b_j(o_t)\}$, where o_t denotes an observation at time t.
- The initial state distribution vector $\pi = \{\pi_i\}$,

  $$\pi_i = P[q_0 = S_i], \quad 1 \leq i \leq N$$

To indicate the complete parameter set of the HMM, the model can be written as follows:

$$\lambda = (A, B, \pi)$$

The information about recognized actions can be further used to validate the performed action and based on this verification either adequate feedback can be generated or the next step can be forced. For example, it is recognized that the user grasped tool number 1, but he should have taken tool number 2, and so he get's an error feedback (e.g. error alert). The verification of the action is done in the *Action Verification component*. This component utilizes the result of the Action Recognition component and the workflow description to check if the right action is performed. Depending on the result of the verification the next step is forced or feedback is generated (e.g. error feedback, corrective feedback).

There is not always a fixed order in which the single steps of a training task have to be passed. In many cases the workflow includes some conditional branches which lead all to the designated goal of the task. For example, if the task is to disassemble a bicycle, one can start with dismounting the front wheel, the back wheel, the handlebar or the seat. Hence, the decision of the next step depends on what is already done. The decision finding is based on the history of performed actions and on the workflow description. Using this information, the most probable next step is selected.

In the ideal case, the workflow block offers the possibility to train the system, or rather to refine the workflow description and action templates using data captured from experts performing the training task. This way, a very natural description of the workflow and the action templates can be achieved, since the resulting description is based on real performances of the task. The workflow block takes data from the capturing controller (while an expert is performing the task) and the workflow description and action templates from the data repository. The capturing data is used as input to train the workflow description, or rather workflow model, and the action templates. The trained models for workflow and actions are sent back to the data repository. One approach to implement this process is the use of Hidden Markov models, as they can not only be used to model and recognize short motion patterns, but also for modeling more complex workflows (i.e. entire tasks) taking into account different additional features, such as a applied forces or contact information [WSB09]. The initial model can be updated incrementally using training samples captured from experts performing the tasks. That is, the model parameters are adjusted to maximize the probability of the observation sequences of the training samples. It has to be taken into account that a human is featured with a large amount of sensors, which are activated when performing a tasks. Since it is nearly impossible to consider all those sensors when capturing the tasks, it is important to choose significant features to be measured. The task pattern that should be recognized (i.e. the captured task performance of the trainee) is then matched with the trained workflow model for example by calculating the likelihoods of the input patterns.

5.4.3.2. Feedback Generation

The interaction processing and application module also generates context information that is presented to the user, such as information about the current progress in the whole task. Moreover, further data referring to the current step is taken from the data repository and used for generating additional information for the user. This data includes for example speech instructions, spatial sound, templates for haptic hints such as rotational or translational movement hints, as well as indications if and how haptic guidance shall be applied. With respect to the current training phase, the accordant representation of the data is selected and sent to the rendering controller.

If a performance error is reported by the workflow block, the interaction processing and application module generates the error feedback, such as an error alert or a vibration stimulus using the template stored in the data repository. If additionally the deviation of the user's performance to the reference data is provided by the workflow block, accordant to the current training phase a corrective feedback is produced. The generated feedback is reported to the rendering module.

Generally speaking, the interaction processing and application module handles all operations for producing output resulting from the user's captured performance and active input, such as the generation of feedback, the control of the visibility of information, or the logging of information about the user's actions and performance.

5.4.4. Multimodal Rendering Controller

The multimodal rendering module handles the rendering of the visual, haptic, and audio data. It is connected with the interaction processing and application module, which delivers the rendering data (see Figure 5.17). The rendering controller contains the necessary interfaces for accessing the rendering systems and devices. It is able configure the rendering systems and devices (if possible) and to feed data into these systems and devices. So for example, it contains interfaces to feed data into a visual rendering system, to control a robot, to control a vibrotactile device, or to feed and configure an audio rendering system.

The rendering techniques to be applied are defined in the system configuration. The depth maps delivered by the capturing module are used for detecting occlusions between real and virtual objects. According to the given configuration, the rendering controller configures the individual rendering by setting all necessary configuration parameters.

The rendering controller also controls the different rendering loops for visual, audio, and haptic rendering and provides parallelized communication channels. The visual, audio, and haptic data is distributed to the corresponding rendering systems, rendered, and presented to the user via display devices.

5.5. Summary

In this chapter, it has been examined how multimodal feedback and Augmented Reality technologies can be applied for training. It has been shown, that especially vibrotactile feedback can be used to enhance AR-based training. In addition, a visualization concept for presenting information in Augmented Reality training environments has been provided as well as a concept for the design of a multimodal Augmented Reality-based training platform that serves for different application domains.

Under consideration of these results, a training concept specific to the training of maintenance and assembly skills using a multimodal Augmented Reality-based platform will be developed in the next chapter.

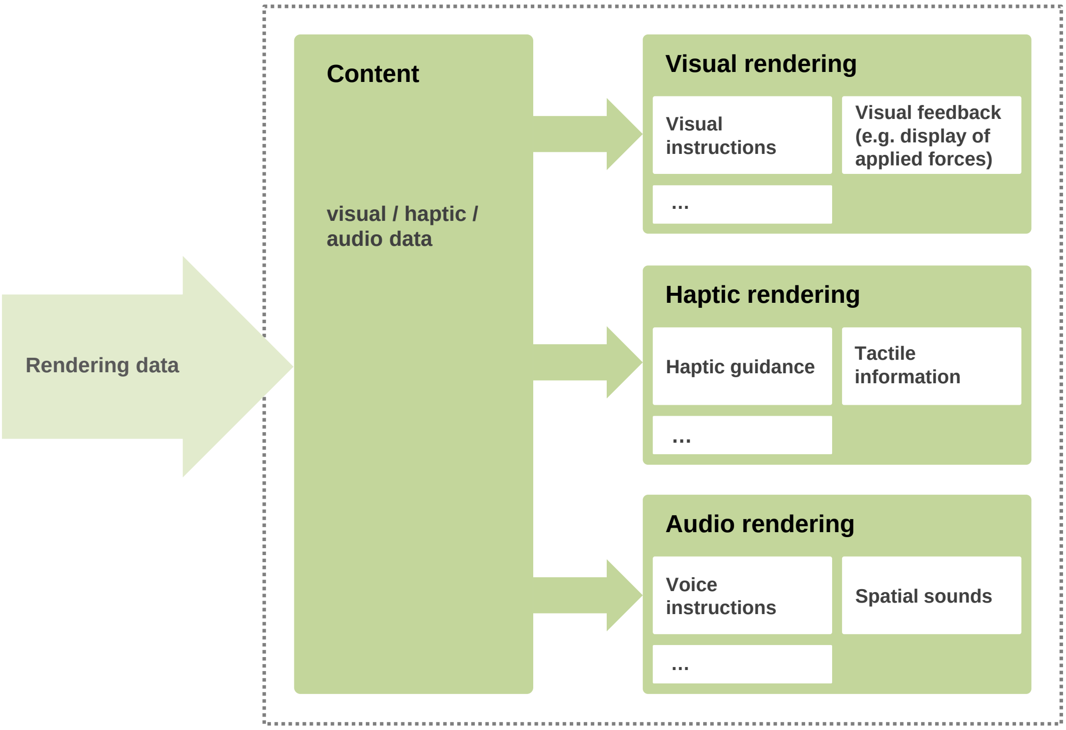


Figure 5.17.: Overview of the multimodal rendering controller.

6. AR-Based Training of Maintenance and Assembly Skills

For the development of an efficient training concept for maintenance and assembly skills, the identification of the underlying sub-skills that are relevant for performing and acquiring maintenance operations is vitally important. Furthermore, it must be examined how these skills can be trained efficiently. This chapter addresses the identification of sub-skills that are involved in the performance and acquisition of maintenance tasks as well as the training of maintenance skills having regard to the most relevant sub-skills.

6.1. Identification of Necessary Sub-Skills

Industrial maintenance and assembly can be considered as a collection of complex tasks. In most cases, these tasks involve the knowledge of specific procedures and techniques for each machine. Each technique and procedure requires cognitive memory and knowledge of the way the task should be performed as well as fine motor "knowledge" about the precise movements and forces that should be applied. Hence, the maintenance skill, which is responsible for a fast and robust acquisition of maintenance procedures, is a complex skill that is composed of different primary skills. In order to develop an efficient system for training maintenance skills, in a first step the most relevant sub-skills involved in maintenance procedures must be identified. These sub-skills comprise sensorimotor skills as well as cognitive skills.

6.1.1. Sensorimotor Sub-Skills

The primary sensorimotor skills that are involved in most maintenance and assembly procedures are listed here:

- *Fine control:* The fine control of hand/finger movements and force applications is an important aspect for performing maintenance procedures. When the technician has to manipulate delicate parts or when the accessibility of the parts to manipulate is low, he must be able to precisely control his finger and hand movements. Especially in cases which include the potential danger of damaging the part of interest with improper movement or applying inappropriate force, the fine control of his hands and finger movements as well as the fine control of his force application to the object are vitally important.

- *Bi-manual coordination:* Some maintenenace task require the use and coordination of both hands working together to perform the task successfully. For example, bi-manual coordination is required, when the technician has to manage a piece and a tool, like he is doing, for instance, when holding a machine part in one hand and loosening a screw at this part with a screwdriver he is holding in the other hand. Therefor he need to coordinate both hands working together at one task.
- *Balance and posture control:* When the accessibility of the task scenario is very low, the technician sometimes has to adopt postures which do not correspond to natural postures, or which are not optimal to undertake the task. In this case, a good balance and posture control is required. For example, if a technician has to move a lever at the side of machine where the accessibility is low, he may have to lean forward to reach the lever. In order to be able to keep this position during the performance of the task, the technician's ability to control his balance and posture is an important aspect.

Fine control is the most relevant sensorimotor sub-skills for maintenance and assembly procedures, since it is vitally important for nearly every maintenance and assembly task. Those tasks usually require a comparatively simple bi-manual coordination that is in many cases not totally new to the technician (e.g. loosening screws from something holding in the hands). So the bi-manual coordination should be considered in the training, but it should not be in focus of training. The same applies for the balance and posture control, which is only relevant for special situations and is not a crucial factor for every maintenance or assembly task.

6.1.2. Cognitive Sub-Skills

As mentioned, there are also cognitive sub-skills involved in maintenance and assembly tasks. The more relevant of those sub-skills are described here:

- *Procedural skill:* In maintenance and assembly tasks, procedural skills reflect the operator's ability to obtain a good representation of task organization, what should be done in each step, and how it should be done (what-when-how). For example, they reflect how to efficiently perform the necessary steps in assembly/disassembly or in removing a broken part of a machine. The technician's ability to identify the necessary operations to execute in order to reach the designated goal is essential for successfully performing the task. For this reason, procedural skills are considered as the most important skills in maintenance and assembly tasks. Later in this chapter procedural skills will be described in detail.
- *Perceptual-observational skill:* This skill is complementary to the procedural skill. Knowing what to do in each step, when to do it, and how to do it, can be based not only on procedural memory, but also on a good and correct mental model of the machine to be manipulated. Being able to identify the elements of a machine and the way it is composed can help the operator to perform the assembly/disassembly task even without consciously memorizing the procedures. In the performance of a frequently performed maintenance and assembly task, the operator is assumed to apply both procedural knowledge and perceptual-observational skills, integrated in

various formats. How a person perceives the environment is not only based on purely cognitive aspects, the perception is also influenced by his interaction with the environment (e.g. manipulation of objects). Hence, perceptual-observational skills represent a combined and integrated cognitive and sensorimotor experience.

The following cognitive sub-skills may also be relevant for maintenance and assembly tasks:

- *Coping strategies and response schemata:* During a maintenance procedure the technician, for example, needs to decide about the necessary operation when detecting a broken part, to select among different strategies to perform the maintenance tasks according to the circumstances, to select between the possible sequences of actions to perform the tasks, or to select the most suitable tools for the tasks. Therefor he needs to access coping strategies and response schemata.
- *Control flexibility and attention management skills:* A certain level of flexibility to cope with unforeseen events that may arise during the task execution is also helpful for performing maintenance tasks. For example, when a specific tool is lacking, or when some parts are missing during the assembly, the technician should be able to cope with these problems and decide the next step.
- *Memory organization, structure and development of knowledge schemata:* During the execution of the assembly and maintenance tasks the technician should be capable of organizing the sequence of sub-tasks and what should be done in each sub-task in order to be able to accomplish the task within a reasonable time. He should be able to encode plans and procedures, knowledge about the task and about the machine (e.g. how to detect a broken part, how to replace it, which tools to use), and to retrieve this information about the necessary sequences of actions to perform when the the task is executed at a later date.

6.2. Training of Maintenance and Assembly Skills

6.2.1. Training of the Required Skills

It has been pointed out that training of humans can be basically split up into three components: training of cognitive, perceptual, and motor demands (see e.g. [PRWEH98]). The cognitive demands concern primarily the construction of an internal model of the task within the trainee's memory. During training the trainee learns gradually how the single parts of a task fit together and where they fit within the final model, and he develops a strategy for building the final model. In the context of maintenance and assembly tasks, the trainee learns how to put the machine parts together, where they fit within the machine, and what steps he has to perform in order to successfully maintain or assemble/disassemble the machine. At the beginning of the training he presumably make mistakes, but the number of mistakes decreases during training, and in later stages he can perform the work correctly and without hesitation.

The training of motor demands concerns the physical performance of the task and the physical handling of all objects involved in the task. The trainee must learn how to deftly cope with the task and the involved objects. That is, he has to learn how to deftly manipulate the machine parts, how to handle them, how to move them within the environment, how to orient them, and how to connect them together to complete the task. This is an important aspect, since a poor training of motor demands leads to misaligned or damaged machine parts (e.g. if they are dropped) and a slow progress in the task execution.

The main objective of the perceptual component of training is to bind the cognitive and motor aspects in order to push the trainee's comprehension of the environment and of the objects involved in the task. The trainee must learn to classify the objects in the environment depending on their characteristics that he perceives (e.g. shape, size, rigidity, friction). This points out that the training of maintenance and assembly skills should address both sensorimotor and cognitive sub-skills.

As described before, three main sub-skills appear to be most relevant for the training. One is sensorimotor and the other two are primarily cognitive. The sensorimotor skill is the fine control of hand movements and force application. The cognitive skills are procedural skills and perceptual-observational skills. The latter one, in essence, represent a combined and integrated cognitive and sensorimotor experience.

Sensorimotor skills can only be acquired through intensive practice in performing the task. A great benefit of using Augmented Reality training environments is the fact, that training can take place in the real environment, or at least in training labs where the trainee can interact with real, physical machine parts and real, physical tools, while receiving supporting virtual information. Maintenance and assembly procedures usually do not require highly complex motor actions which are completely new to the technicians. Technicians have practiced those actions a lot in their daily work, and hence they have already acquired a certain level of the underlying sensorimotor skills. When those technicians are trained in performing new maintenance tasks, the training of the sensorimotor skills should not be the main focus, since they are trained anyway if the training takes place with real objects. Nonetheless, the training should include elements to support the technician in refining his sensorimotor skills, particularly the fine motor control. Such elements should, for example, provide information about forces to apply, give feedback about applied forces, or procure haptic hints for guiding the trainees fine motor trajectories.

The training should rather focus on the cognitive skills which are crucial for acquiring the skill to transfer the trained task to other situations and to accelerate the acquisition of new maintenance tasks. Many studies showed that procedural skills develop gradually as a result of practice through repeated performances of a given task (e.g. [GC02], [And82]). A distinction must be drawn between procedural skills in pure cognitive tasks (e.g. mathematic, language), which can be trained cognitively, and procedural skills in psychomotor tasks, which involve a rich sensorimotor interaction and experience and therefore are mostly developed by performing the task using different modalities (vision, sense of touch, etc.). It is obvious that the latter case applies for training of maintenance and assembly skills. This confirms the statement that the training of cognitive and sensorimotor sub-skills should not be separated. Training the perceptual-observational skills means enabling the technician to create a more accurate and useful representation and diagnostic of machine states and of the actions that must be taken for its maintenance. The use of multimodal feedback can help in achieving this goal (e.g. the visualization of a 3D model of grouped objects forming a logical unit or adding haptic hints may help the trainee in developing accurate mental models).

Due to these reasons, procedural skills, and with this also the perceptual-observational skills, should be in the main focus of the training of maintenance and assembly skills. Important aspects for training procedural skills will be described in the following chapter.

6.2.2. Use of Multimodal Interfaces

As mentioned in Chapter 5, the use of multimodal information channels can improve training enormously. The presentation of information using different modalities can support the trainee's comprehension of instructions about the steps to execute, can support him in performing the steps and in developing internal representations of the task, and can lead to a better distribution of the cognitive work load.

In Augmented Reality-based training of maintenance and assembly skills, multimodal elements can be applied for different purposes. For example, to support the technician in training the fine control of force application, it is beneficial to provide information about the force and/or torque he has to apply and the force and/or torque he is currently applying. Indeed, the value of force and torque to apply can be presented to the trainee using visual objects like text or indicator bars, but for the trainee the displayed values may be difficult to rate and to translate into action. Presenting these values using haptic information like vibrotactile stimuli, they are easier to understand for the trainee: a low intensity of stimuli corresponds to a small amount of force to apply, intensive stimuli correspond to high force application. In contrast, for feedback about the captured forces and torques that the trainee is currently applying a haptic presentation is not suitable and would rather confuse the trainee. In this case, a visual presentation, such as text along with color-coded indicator bars, is much more effective. The measuring of applied forces can be realized by using force/torque measuring tools. One example for such a tool is described in Chapter 4.3.4.

Also the capturing of the force profile can be useful in specific maintenance task. For example, if a machine part needs to have a special temperature before it can be build in, a force profile capturing tool (see Chapter 4.3.4) can be used to measure, if the temperature is adequate when the trainee grasps the machine part. In case the temperature is not adequate, an error feedback (visual, auditory or haptic) can be presented to the user. If also a special force application must not be exceed, the force capturing tool can measure also the force applied by the trainee when assembling the part, and based on this information corrective feedback (e.g. vibrotactile stimuli) can be generated.

The tracking and recognition of hands and objects, such as tools and machine parts, is also useful for training of maintenance and assembly skills, even though this is still a challenging field for unprepared environments. Using this technologies, it can be, for example, determined if the trainee grasps the right machine parts, uses the right tools, or performs the right actions. If not, error feedback can be generated.

The main advantages of using different modalities for presenting information to trainees have already been described in Chapter 5.1.1. It becomes obvious that presenting all information and feedback to the trainee using one modality, usually vision, is not recommendable. On the one hand, this may lead to an overload of visual information and it may be difficult for the trainee to process all this information. On the other hand, vision is not always the most suitable modality for presenting information (cf. example above: providing information about forces to apply). Also for presenting error feedback it can be more effective to use other information channels like the haptic or auditory channel. If an adequate device is available, vibrotactile stimuli can be presented directly at the point of interest (e.g. arm). An audio alert can be quickly recognized by the user and, if spatially arranged, can be provided at the target location.

Apart from all that, it must be considered that the utilized haptic devices must not hinder the trainee in performing the task in a natural way. That is, the trainee must not be restricted in his natural movements. If this would be the case, the training of the sensorimotor skills required for the natural performance of the task would be impaired. For this reason, the use of lightweight vibrotactile devices for providing haptic information is suggested, since those devices can provide useful information about motion patterns to perform (rotational, translational hints), forces to apply, or performance errors.

Depending on the training environment, the presentation of audio information may be limited. The machines to maintain are typically installed in very noisy assembly halls or machine halls, and the technicians have to wear helmets with integrated noise protection. If the training takes place "on-the-job" in this environments, audio information can only be provided by using head-phones that must be integrated in the helmets. Due to security reasons, this is not always allowed (e.g. because of possible audio warning signals in the machine hall).

6.2.3. Use of a Vibrotactile Bracelet

In maintenance and assembly tasks the operator acts mainly with his hands. Thus, the application of haptic feedback to the trainee's hands during the training of those tasks can support him in comprehending how to perform the task and in performing it. Devices like vibrotactile bracelets offer the possibility to apply vibration stimuli to the human arm, forearm, and wrist. Such a bracelet is shown in Figure 6.1. The bracelet is equipped with six vibration actuators which are placed at equal distance

Figure 6.1.: A vibrotactile bracelet developed by DLR (German Aerospace Center) [SEWP10].

from each other inside the bracelet and thus also around the user's arm. The intensity of each actuator can be controlled individually. That way, various sensations can be generated such as rotational or translational movements hints. Weber et al. present a study in which the described bracelet is used for spatial guidance [WSH*11]. The participants, who were wearing the bracelet at their right wrist, had to translate and rotate virtual objects according to the vibrotactile hints without any additional visual information. The outcome was compared to the performance of participants executing the same tasks solely under verbal guidance. By comparing task completion times, movement accuracy, and subjective ratings it was shown that the participants using the vibrotactile bracelet achieved similar performance for translational tasks and better performance for rotational tasks than the participants working under verbal guidance. From this it can be concluded that the use of the bracelet is valuable for guiding simple arm movements, in particular in noisy environments where audio guidance is critical.

Transferring this results into the context of maintenance and assembly training, it can be assumed that providing vibrotactile feedback using a bracelet is useful for presenting movement hints for maintenance tasks. The bracelet can be used to guide the trainee to the target object he has to manipulate. Furthermore it can indicate movements which the trainee needs to perform, such as rotational movements for unscrewing objects or translational movements for moving a lever. Beside giving movement hints, it can also be used to provide error feedback (e.g. a short impulse in case of error such as grasping the wrong tool) or to describe the action the trainee has to perform (e.g. knocking-like impulses for describing knocking-like action like hammering). In addition, it can approximately indicate the intensity with which actions must be performed (e.g. strong stimuli indicates high effort, soft stimuli indicate delicate actions).

6.3. Training of Procedural Skills

6.3.1. Development of Procedural Skills

Procedural skills are the ability to follow repeatedly a set of actions step-by-step in order to achieve a specified goal, and they reflect the operator's ability to obtain a good representation of task organization. This skill is needed in the performance of complex tasks as well as of simple tasks. As described in Chapter 6.1.2, procedural skills are the most important skills for maintenance and assembly tasks.

Procedural skills are based on two main components: procedural knowledge and procedural memory. Procedural knowledge enables a person to reproduce trained behavior. It is defined as the knowledge about how and when (i.e. in which order) to execute a sequence of procedures required to accomplish a particular task [And82]. Procedural knowledge is stored in the procedural memory, which enables persons to preserve the learned connection between stimuli and responses and to response adaptively to the environment [Tul85].

According to Anderson's three-stage model described in Chapter 3.2.1, the acquisition of cognitive skills, and thus also of procedural skills, goes through three major stages. In the declarative stage the learner receives instructions and information about the skill and encodes this information as a set of facts about the skill. During the knowledge compilation process, with practice the knowledge is converted into a procedural form in which it is directly applied without the intercession of other interpretive procedures [And82]. It has also been suggested that procedural learning develops through incremental tuning of the processing of elements that underlie a dedicated task [Coh84]. Another theory points out that procedural learning is an incremental process of transduction between representations (performed through transducers) [GC02]. Each time the task is performed, the transducers that are involved in the task performance are tuned. The improvement of the effectiveness of the transducers, which results from their incremental tuning in repeated performances of the task, leads to the improvement of performance.

To summarize, procedural skills develop gradually over several sessions of practice (e.g. [GC02]) and are based on getting a good internal representation of a task organization: what appropriate actions should be done, when to do them (appropriate time) and how to do them (appropriate method). Therefore, the training of procedural skills should address the development of a good internal representation of the task and the execution of the single steps in the right order in early training phases.

6.3.2. Enhancement of Mental Model Building

It has been explored that the performance of a learner of a procedural skill becomes more accurate, faster, and more flexible when he is provided with elaborated knowledge ([FOO93], [APH*03], [THDA08]). This means that the learner's performance increases when *how-it-works knowledge* (also called *context procedures*) ([Kie88], [FOO93], [THDA08]) is provided in addition to the *how-to-do-it knowledge* (also called *list procedures*). According to Taatgen et al., the learner is able to extract representations of the system and the task which are closer to his internal representation when elaborated knowledge is given, and as a result his performance improves [THDA08]. This internal, psychological representation of the device to interact with can be defined as *mental model* [AQP*01].

In order to support the trainee's mental model building process, the features of the task which are most important for developing a good internal representation must be presented to the trainee. It has been suggested that "the mental model of a device is formed largely by interpreting its perceived actions and its visible structure" [Nor87]. The mental model building is mainly influenced by two factors: the actions of the system (i.e. the task and the involved device) and its visible structure.

Transferring this into the context of procedural skill training, two aspects seem to be important for supporting the development of a good mental model: One is providing an abstract representation of the system by what a better understanding of how it works is constructed. The other aspect is providing the visual representation of the system, which will strengthen the internal visual image.

When the device, or rather the assembly procedure, is complex, it is better to present the user only those sub-parts of it which are relevant for the current step the user has to perform, instead of presenting the entire model ([NM00], [APH*03]). Furthermore, people think of assemblies as a hierarchy of parts, where parts are grouped by different functions (e.g. the legs of a chair) [APH*03]. Hence, the hypothesis is that the displayed sub-part of the assembly task should include both the condition of the device before the current step (or rather the logical group of steps to which the current step belongs) and the condition after. This hypothesis is based on the work of Taatgen et al. [THDA08], in which it is shown that instructions stating pre- and post-conditions yield better performance than instructions not doing so. Reviewing this it can be concluded that the user's mental model building process can be improved by using a visualization providing context information.

To decide which steps of the task form a logical unit (e.g. all legs of a chair) it is important to obtain information from experts, or rather to capture their mental model. One possibility to capture the mental model is to extract the required information from verbal or written reports of experts, such as interviews, questionnaires, or "think aloud" protocols (an expert thinks aloud while performing the task "in mind"). Another possibility is to analyze empirical performance measures (e.g. the sequence of actions that an expert performs, the time he needs to initiate an action, etc.).

6.4. Development of the Training Program

In the following the development of a training program for the training of maintenance and assembly skills and the evaluation of the used training system is described. This includes the definition of accelerators and the design of the training protocol.

6.4.1. Specific Accelerators

An important factor in the development of the training program is the definition of accelerators which are introduced in the training to facilitate and improve learning. In the following accelerators for the training of maintenance and assembly skills will be suggested and described.

In order to enhance the training of maintenance and assembly skills we propose to pursue the following training strategies:

- The provision of *direct aids*
- The provision of *indirect aids*
- The provision of *context aids*

Direct aids are aids that are permanently present (i.e. the user does not have to explicitly request them) and that provide "direct" information about the actions to perform (e.g. a direct spatial information of where to perform the action or a superimposed 3D animation showing the action to perform). *Indirect aids* pursue two aims: first, they enhance the information delivered by direct aids by providing further details (i.e. detailed instructions) on user demand or by providing abstract and subliminal information that the user needs to interpret and transfer to the task, and secondly, they provide information concerning the complete step to perform and the involved components (not only about the action), such as hints about the required tools, screws or washers. Since indirect aids do not strongly guide the user, they do not avoid the user's active exploration of the task, which is important for learning (see Chapter 3.3). *Context aids* present information about the context of the step to perform. That is, they provide hits about the structure of the task (and the device) and about the user's progress in the task execution in order to support the user's mental model building process (see Chapter 6.3.2).

The definition of concrete accelerators realizing and implementing these training strategies and is presented in the following. The evaluation of the accelerators will be presented later in this work (see Chapter 8).

6.4.1.1. Adaptive Visual Aids with Information on Request

As described in Chapter 3.3, guided experience is good for learning, but an active exploration of the task has to be assured as well. A too strong guidance of the trainee during training impedes an active task exploration and harms the learning process. Active exploration naturally occurs when transferring the information about the task during training is accompanied with some difficulties, forcing the trainee to independently explore the task. If such difficulties are reduced (e.g. by showing the user in detail how to solve the problem), active exploration may not take place. Strong visual guidance tools impede active exploration, because they guide the trainee in specific actions and thus inhibit the trainee's active exploratory responses [YGY10]. This can be illustrated using the example of a car driver guided by a route guidance system: This driver typically has less orientation than driver who is exploring the way with the help of maps and street signs. Also reproducing the way when he has to drive it again is more difficult for the driver who used the route guidance system. From all this it can be concluded that the training process must include visual aids which allow for reducing the level of displayed information (i.e. flexible visual aids), in order not to avoid the trainee's active exploration of the task by providing a too strong guidance.

Another important aspect is, how much information should be visualized in the different training phases. A basic understanding of how much information the trainee needs during learning can be obtained by observing studying people. Examining the learning behavior of a student studying procedural processes using textbooks or written notations, the following characteristics can be observed: First of all, for each step the student marks a couple of words, a sentence, or an excerpt in the running text and writes annotations at the side margin. He studies the process by going repeatedly through this learning material. In the first cycles, the student reads the marked text and the accordant annotations to catch information about the single steps and to put them in order. With the increasing number of performed studying cycles the information that he needs to decide and reproduce the single steps of the procedure decreases. That is to say, at first he needs all information available to reproduce the steps. After some cycles he uses only the marked text to remember the steps, and later he highlights only buzzwords in the marked text and uses them as hints for reproducing the next step. Anyhow, whenever he needs more information to reproduce the step, he can "ask for it" by looking at the annotations on the side margin. It becomes apparent that the student gradually decreases the information he uses for reproducing the steps and their order. When he starts studying he needs more detailed information about the single steps, because the learning of the single steps is in focus. With the growing development of an understanding of the single steps, the learning of how the steps fit together (i.e. of the procedure) comes increasingly to the fore. The observed behavior corresponds to the development process of procedural skills described in the previous chapter.

By transferring this observation into the context of training, the mapping of the visualized information level to the different training phases can be hypothesized as follows: In early phases, a clear and detailed instruction about the current step should be provided in order to train the trainee in understanding and performing the single steps. This can be realized by using *Adaptive Visual Aids* consisting of overlaid 3D animations (pointer) and/or multimedia instructions (content). Alternatively, the pointer can act as object/area highlight while the content provides the detailed multimedia instruction. During the training the level of presented information should be gradually reduced (e.g. only 3D animation, then only area highlight with some buzzwords or a picture, then only area highlight, etc.). In phases in which no visual aid content is directly visualized, it should be possible for the trainee to ask for the last available content (i.e. the last displayed content of a elapsed training phase). This is particularly important, if the trainee cannot perform the step without the additional information. In this case, if he could not get the additional information, he would not be able to continue the task.

6.4.1.2. Device Display

It has been discussed that the presentation of context information, such as logical units of sub-tasks, and the display of the device to maintain (e.g. to assemble/disassemble) can support the trainee's mental model building. Moreover, the presentation of only relevant sub-parts of the device and the visualization of the pre- and post-conditions can further enhance the development of a good internal representation (see Chapter 6.3.2).

Based on these findings we suggest the use of a "device display". The device display is a visual element that provides information about successive steps, or rather sub-tasks, which belong to a logical group. That is, it provides information about a good mental model of the task. This supports the user in developing his internal representation of the task. The provided information includes also the condition of the device before the current step and afterwards. Thus, using the device display, the user can recognize a sub-goal of the task he has to perform. This can help him to understand "what" he has to do, and hence to deduce the next step to perform. In fact, the presentation of sub-goals actually forces the trainee to deduce the next step without using a more direct visual guidance.

The device display can be realized as a view-aligned object containing a video or a 3D animation showing the grouped sub-tasks and the corresponding state of the device. The visibility of the device display can be toggled on user demand. Examples for the visualization of a device display are shown in Figure 6.2.

6.4.1.3. Structure and Progress Information

Providing abstract, structural information about the task can improve the trainee's mental model building process, and hence the acquisition of procedural skills (see Chapter 6.3.2). Therefore, visual elements presenting information about the structure of the training task should be included in the training system. Not only the structure of the task, but also the relation between the current status and the structure is important. That is, the position of the current state in the whole structure should be visualized

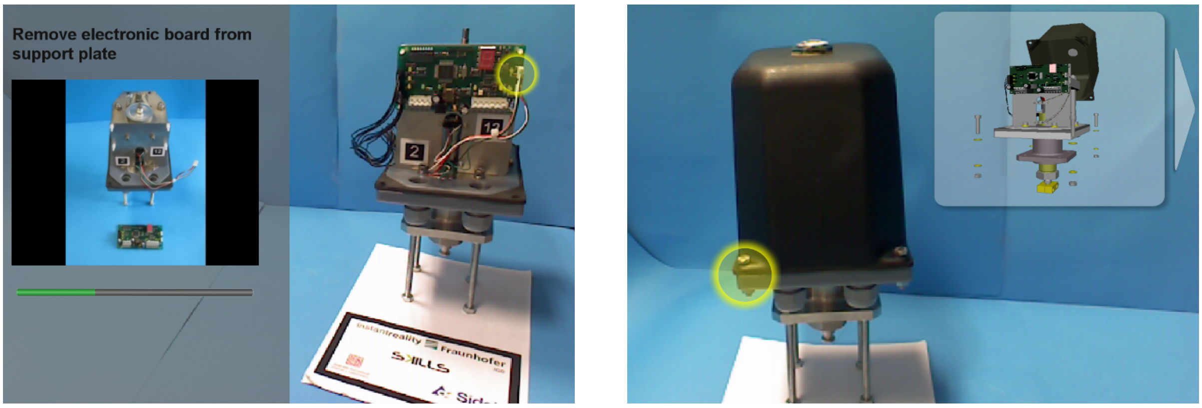


Figure 6.2.: Different realizations for a device display.

as well. Thus, the trainee gets an overview of the training task. He can arrange the current step in the structure of the task and use this information to refine his internal representation of the task.

One possibility to visualize structural information is using progress-bars. Progress-bars provide an abstract overview of the trainee's current status in relation to the whole task. They can also be included in the device display to demonstrate the trainee his progress inside a grouped task (i.e. his progress in achieving a sub-goal).

6.4.1.4. Haptic (Vibrotactile) Hints

The potential of vibrotactile feedback for maintenance and assembly tasks and its use as guidance tool have been demonstrated in Chapters 5.1.3 and 6.2.3. Due to these findings we suggest the integration of haptic or rather vibrotactile hints into the training system. Usually a lot of visual information has to be processed in complex working scenarios. In contrast, the tactile channel is less overloaded. Furthermore, vibrotactile feedback is a quite intuitive feedback, as the stimuli are directly mapped to body coordinates. Since it provides a soft guidance which "channels" the user to the designated target instead of directly manipulating his movements, it does not prevent the active exploration of the task. Thus, the mental model building process can be supported. Vibrotactile hints can be given by using simple devices like the vibrotactile bracelet described in Chapter 6.2.3.

Vibrotactile feedback should be used to give the trainee additional movement hints, such as rotational or translational movements cues, and to guide the trainee to specific targets. For example, if the trainee needs to rotate his arm for performing a sub-task, the rotational direction (clockwise or anti-clockwise) may be difficult to recognize in a video showing an expert performing the sub-task. Receiving the same information using a vibrotactile bracelet, the trainee can easier identify the rotational direction. Also translational movements can be conveyed using vibrotactile stimuli.

Apart from that also for presenting error feedback, such as communicating whether the right action is performed (e.g. the right tool is grasped), it is useful to apply vibrotactile feedback. This is a significant factor, as it can prevent the user from performing errors at an early stage. In addition, vibrotactile hints should be used to provide slight instructions by directing the trainee's attention to a body part and indicating movements (e.g. rotational movements).

6.4.2. Training Protocol

In the following a training protocol for an exemplary real world maintenance task scenario will be described. This task scenario has been implemented and used for evaluating the accelerators and the training system.

6.4.2.1. Definition of Task Scenarios

The exemplary maintenance task consists of replacing the belt roller, the clamp and the level sensors of the actuator shown in Figure 6.3. This actuator is a component of a filling station. During training, the trainee learns the procedures that are necessary to efficiently replace these components without damaging any part of the actuator. For this purpose the actuator and its support plate holding the two sensors must be disassembled and reassembled after replacing the damaged components. Therefore, the training program contains two different task scenarios: disassembly and assembly. In the first scenario, the actuator must be disassembled in order to remove the sensors from the support plate and to dismount the belt roller and the clamp. This disassembly task is assessed as successfully performed, if the all mentioned machine parts have been disassembled without the occurrence of any damage to the actuator and its components. In the second scenario the sensors, the belt roller, the clamp, and the actuator must be reassembled. Similar to the disassembly procedure, the performance of this assembly task is rated as successful if the replaced components and the actuator have been assembled without the occurrence of any damage.

To disassemble the actuator and to dismount the belt roller and the clamp and the sensors five major subtasks must be performed. These subtasks have been defined by expert technicians and trainers from industry.

Sub-task A: Remove the actuator cover

Sub-task B: Dismount the circuit board

Sub-task C: Dismount the support plate and the two level sensors

Sub-task D: Dismount the belt roller

Sub-task E: Remove the clamp from the actuator stem

Each of these sub-tasks requires the execution of a number of single steps. These steps are listed below.

A: Remove the actuator cover: two steps

- **A-1:** Unscrew the four fixing screws of the actuator cover

Figure 6.3.: The selected exemplary training scenario consist of replacing two sensors and the belt roller of this actuator. For this purpose the actuator must be party disassembled and reassembled after replacing the damaged components.

- **A-2:** Remove the actuator cover

B: Dismount the circuit board: four steps

- **B-1:** Disconnect the drive cable from the circuit board
- **B-2:** Cut the holding clip that keeps the sensor cables together
- **B-3:** Disconnect the two sensor cables from the circuit board
- **B-4:** Remove the circuit board from the support plate

C: Dismount the support plate and the two level sensors: four steps

- **C-1:** Remove the fixing screws of the support plate
- **C-2:** Remove the support plate from the structure of the actuator
- **C-3:** Unscrew the screws fixing the lower sensor to the support plate and remove the sensor from the support plate

C-4: Unscrew the screws fixing the upper sensor to the support plate and remove the sensor from the support plate

D: Dismount the belt roller: three steps

D-1: Remove the nut that holds the belt roller in place

D-2: Remove the belt roller with the belt

D-3: Separate the belt roller and the belt

Remove the clamp from the actuator stem: two steps

E-1: Unscrew the two fixing screws of the clamp

E-2: Remove the clamp

Consequently, the disassembly scenario consists of fifteen single steps. The defined sub-tasks form logical units of steps which must be performed in a specific order to reach a sub-goal of the entire task scenario (here: disassembly). Hence, the sub-tasks represent a "mental model" of the task. The order in which the sub-tasks and their underlying single steps have to be performed is also part of the training protocol. Figure 6.4 shows the hierarchical task analysis of the disassembly scenario determined by expert technicians and trainers.

Conversely to the disassembly task, the assembly scenario is composed of the following sub-tasks, which are necessary to perform in order to reassemble the replaced components and the actuator.

Sub-task A: Place and fix the actuator cover

Sub-task B: Mount the circuit board

Sub-task C: Mount the two level sensors and the support plate

Sub-task D: Mount the belt roller

Sub-task E: Mount the clamp to the actuator stem

In analogy to the disassembly procedure, also these sub-tasks form logical units of steps and hence represent a "mental model" of the assembly task. The single steps that underly the sub-tasks are composed as follows.

A: Place and fix the actuator cover: two steps

A-1: Place the actuator cover onto the actuator

A-2: Place and tighten the four screws fixing the cover to the actuator

B: Mount the circuit board: four steps

B-1: Attach the circuit board to the support plate by clicking it onto the spacers of the plate

B-2: Connect the sensor cables to the circuit board

B-3: Fix the sensor cables together using a cable tie

B-4: Connect the drive cable to the circuit board

C: Mount the two level sensors and the support plate: four steps

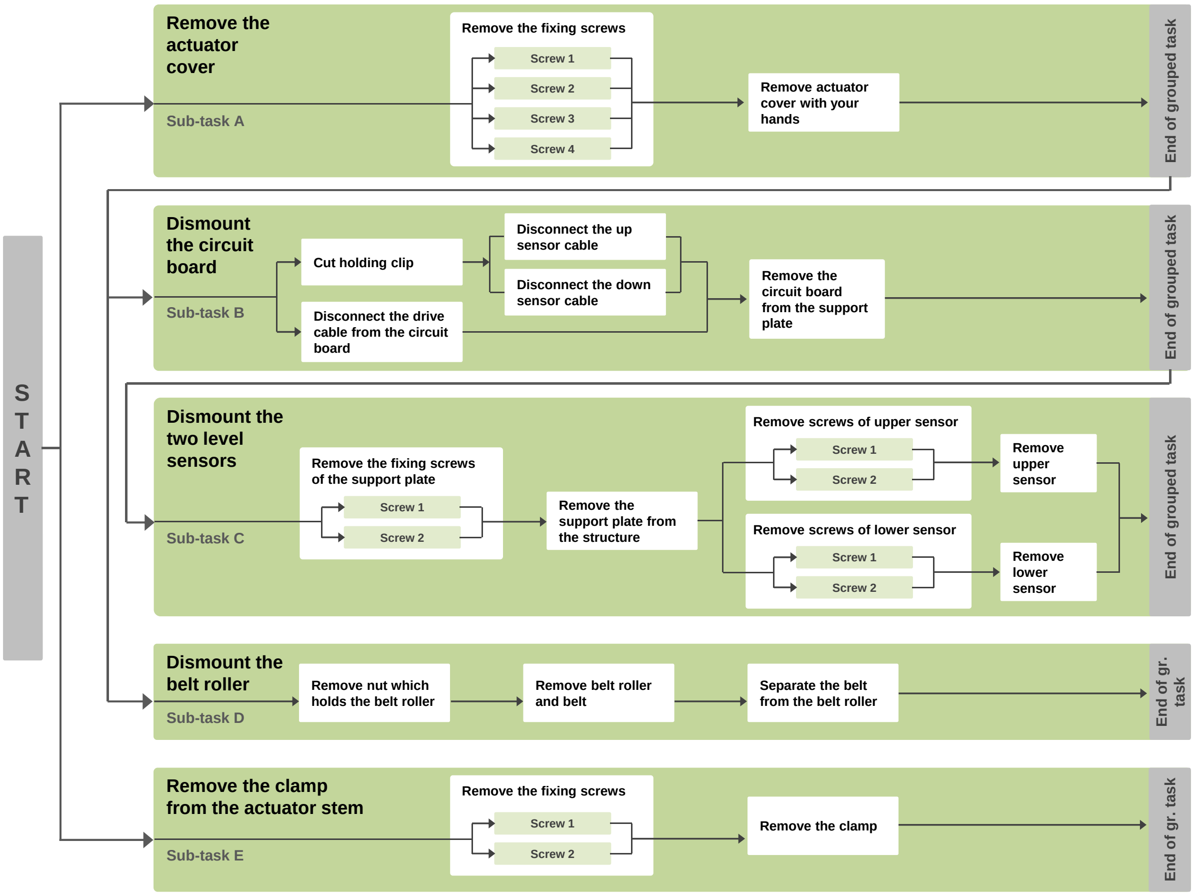


Figure 6.4.: Hierarchical analysis of the disassembly task scenario and its logical units.

C-1: Position the upper sensor and tighten the screws fixing it to the support plate

C-2: Position the lower sensor and tighten the screws fixing it to the support plate

C-3: Attach the support plate to the structure of the actuator

C-4: Place and tighten the two screws fixing the support plate to the structure of the actuator

D: Mount the belt roller: three steps

D-1: Join the belt roller to the belt

D-2: Position the belt roller with the belt

D-3: Place and tighten the nut that holds the belt roller in place

E: Mount the clamp to the actuator stem: two steps

E-1: Position the clamp

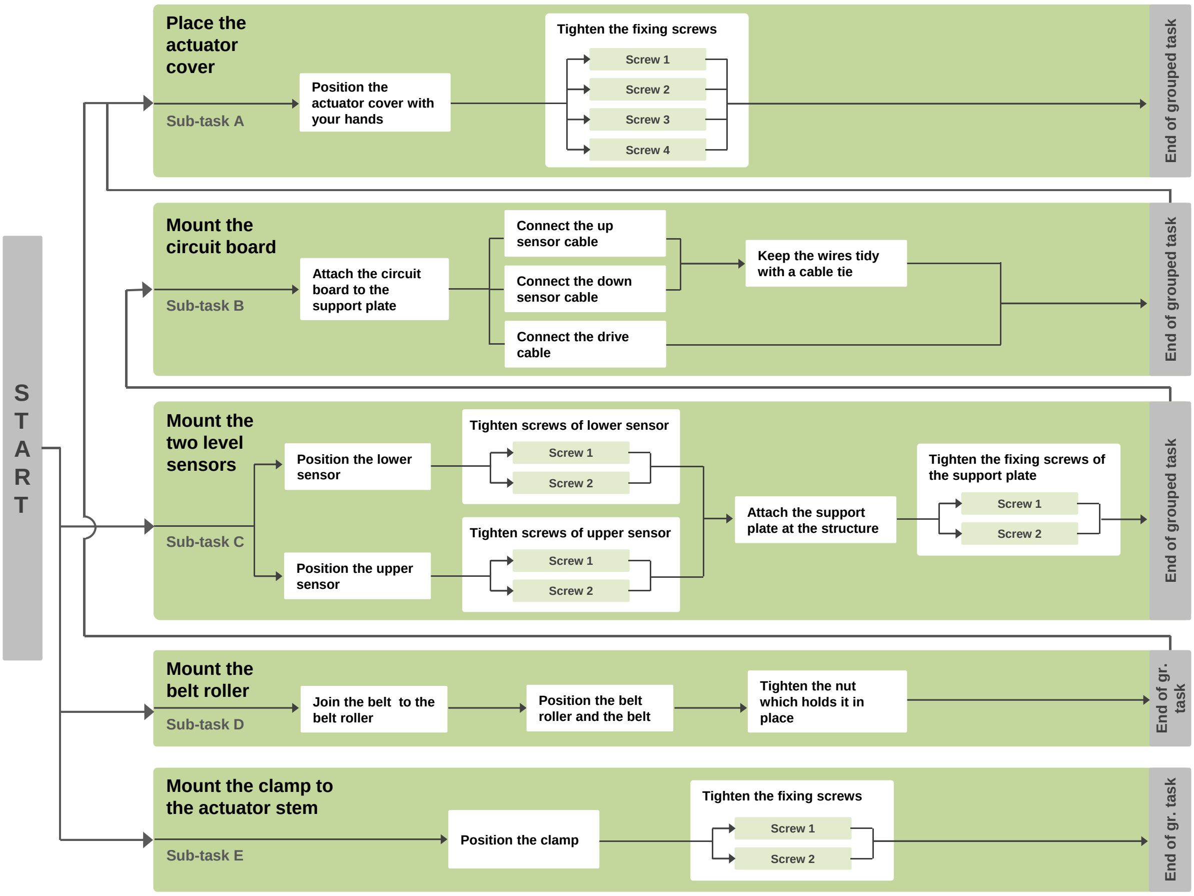


Figure 6.5.: Hierarchical analysis of the assembly task scenario and its logical units.

E-2: Place an tighten the two screws fixing the clamp to the actuator

An overview of the hierarchical analysis of the assembly task consisting of fifteen single steps and the specified logical units is shown in Figure 6.5.

Also the definition of the type and amount of information and feedback provided in different training phases is part of the training protocol. Depending on the training material that is available (text descriptions, instruction videos and images, 3D models and animations, etc.), a mapping of the information to be displayed to the training phases must be specified for each step and each logical unit of the training tasks. An exemplary assignment between provided information and training phases for the selected assembly training task is shown in Table 6.1, Table 6.2, and Table 6.3. An analog mapping for the disassembly task has to be defined in the training protocol as well.

Assembly: Structure/Progress information		
Phase	**Structure/Progress information**	**Devive Display**
1	detailed progress information (entire task and sub-tasks), transparent 3D model (in superimposition with camera image)	yes
2	detailed progress information (entire task and sub-tasks), transparent 3D model (in superimposition with camera image)	yes
3	detailed progress information (entire task and sub-tasks)	yes
4	simple progress information (only entire task)	no

Table 6.1.: Exemplary mapping of provided structure and progress information to the training phases for the assembly task.

6.4.2.2. Training Process

Trainee's who have not used the training platform before learn how to handle the platform, its devices and the training application in a *familiarization session* which takes place before the beginning of the actual training. In the familiarization session the trainee learns how to utilize the devices (e.g. tablet PC, haptic bracelet), how to initialize the tracking, and how to interact with the application. This is done by providing all involved physical devices of the training platform and explanatory documents, such as slides or user manuals, and/or a trainer on-site explaining the trainee the utilization of the platform. That way the trainee can practice the handling of the platform and the application. This is important to allow the trainee for concentrating on the training itself during the training procedure instead of focusing on the handling of the platform and the application.

The actual training takes place in a real industrial training environment. The trainees work with real components and real tools (no dedicated training tools or devices). This way, the training conditions are consistent and very close to the real-world conditions in the machine hall. One training session for a trainee consists of performing one time the complete training scenario (i.e. disassembly and assembly task) using the training system. After each training session the trainee's performance is analyzed to determine his training level and choose the accordant training phase for the next session. The performance measures used for analyzing the trainee's performance are described in the following section. The training should not exceed two sessions per day, where on is conducted in the morning and the other one in the afternoon. In case the training is performed in too close succession, the acquisition of procedural knowledge about the task can not be reliably assessed. The training process should be conducted until the trainee is able to perform the training tasks with a very good performance in the highest training level.

Assembly: Logical units/Sub-tasks		
Sub-task	**Phase**	**Devive Display information**
A	1	short task description, textual information about required tools, 3D animation
A	2	sub-task title, 3D animation
A	3	sub-task title, pre- and post-condition of device (image)
B	1	short task description, textual information about required tools, instruction video
B	2	sub-task title, short instruction video
B	3	sub-task title, pre- and post-condition of device (image)
C	1	short task description, textual information about required tools, instruction video, close-up images of mounted sensors
C	2	sub-task title, short instruction video, close-up images of mounted sensors
C	3	sub-task title, pre- and post-condition of device (image), close-up images of mounted sensors
D	1	short task description, textual information about required tools, instruction video
D	2	sub-task title, short instruction video, close-up image of complicated detail
D	3	sub-task title, pre- and post-condition of device (image)
E	1	short task description, textual information about required tools, instruction video
E	2	sub-task title, short instruction video
E	3	sub-task title, pre- and post-condition of device (image)

Table 6.2.: Exemplary mapping of displayed information to the training phases for the logical units of the assembly task.

Assembly: Single steps				
Step	**Phase**	**AVA pointer**	**AVA content**	**Haptic Hints**
A-1	1	3D animation	task description, instruction video	—
A-1	2	object highlight	short task description, instruction video	—
A-1	3	area highlight	sub-task title, short instruction video	—
A-1	4	area highlight	sub-task title	—
A-2	1	3D animation	task description, instruction video, required tools (text/image), required components such as screws, nuts, machine parts, etc. (text/image)	rotational
A-2	2	object highlight	short task description, instruction video, required tools (text/image), required components (image)	rotational
A-2	3	area highlight	sub-task title, short instruction video, required tools (text/image), required components (image)	rotational
A-2	4	area highlight	sub-task title, required components (image)	—
B-1	1	3D animation	task description, instruction video	—
B-1	2	object highlight	short task description, short instruction video	—
B-1	3	area highlight	sub-task title, pre- and post-condition of device (image)	—
B-1	4	area highlight	sub-task title	—
B-2	1	3D animation	task description, instruction video	—
B-2	2	object highlight	short task description, short instruction video	—
B-2	3	area highlight	sub-task title, pre- and post-condition of device (image)	—
B-2	4	area highlight	sub-task title	—
B-3	1	3D animation	task description, instruction video, required components (image)	—
B-3	2	area highlight	short task description, short instruction video, required components (image)	—
B-3	3	area highlight	sub-task title, pre- and post-condition of device (image), required components (image)	—
B-3	4	area highlight	sub-task title, required components (image)	—
B-4	1	3D animation	task description, instruction video	—
...	...	...	...	...

Table 6.3.: Exemplary mapping of provided information to the training phases for the single steps of the assembly task.

6.4.2.3. Performance Measures

As mentioned, the trainee's performance is assessed after each training session. The performance of carrying out both training tasks is measured using the following criteria:

Completion time of task The time needed by the trainee to accomplish the training task using the training platform.

Completion times of steps The time needed by the trainee to perform the single steps of the training task using the training platform. The completion time of each step is measured.

Requests for aid The number of requests for non-permanently provided aids (e.g. request for AVA content, request for Device Display) during the performance of the training task. Also the steps in which the trainee requested for information is logged.

Performance errors The number and type of errors (e.g. used inadequate tool, forgot washer) in the trainee's performance of the training task. Also the steps in which the errors were made is logged.

The listed performance measures can not be considered completely independent, but rather have to be set in correlation. For example, a short task completion time does not per se indicate a good performance; rather it must be set in relation to the number of performance errors the trainee made when accomplishing the training task and the number of his requests for additional aids. Consequently, a higher completion time of the task is not implicitly an indicator for a poor task performance; if the trainee achieved the goal of the task without making any errors and with only few requests for non-permanently aids, the performance is not necessarily assessed as very poor. Correlating the measured step completion times with the task completion time provides information about the trainee's current level of procedural knowledge about the task. If the sum of all step completion times is much smaller than the time the trainee needed to perform the complete task, he spend much time in trying to decide the next step without the help of the system. Hence, he needed much time to recall knowledge about the task procedure. In case the difference between task completion time and the added up step completion times is minor, the trainee can quickly access and interpret his procedural knowledge about the task, which is an indication for a high level of procedural knowledge.

The use of an adequate scoring considering the different performance measures is an important factor for evaluating the trainee's performance of the training tasks. The weights of the performance measures can not be generally defined, but rather depend on the condition of task and device. For example, in tasks that are composed of a large number of steps which are easy to perform a good ability to recall knowledge about the task structure and procedure (i.e. require good procedural skills) is a key factor. In contrast, tasks consisting of few steps that are complicated to perform require mainly the knowledge about how to accomplish the single steps and the motor abilities to perform them. Hence, for tasks being composed of many simple steps the task completion time—or rather the difference between the task completion time and the sum of step completion times—has a higher weighting compared to the tasks consisting of only few complicated steps.

The collected information about the steps in which the trainee asked for additional help and the steps in which he made errors can not only serve as performance measure, but it can also be used to discover the steps that are complicated to perform and that require good instructions. By means of this data, the information provided in each step can be assessed and if necessary refined in order to improve the support for the trainee during training.

6.5. Summary

In this chapter, a concept for training maintenance and assembly skills using multimodal modalities and Augmented Reality technologies has been presented. This includes the decomposition of maintenance in the underlying sub-skills as well as the identification of the most relevant sub-skills, namely the so-called procedural skills. It has been pointed out that supporting the user's mental model building process is a crucial factor for the training of procedural skills. Training strategies and specific accelerators to support the user during the training and to improve his skill level have been defined. Furthermore, a training program for the training of maintenance and assembly skills has been developed.

In the following chapter, the implementation of a multimodal Augmented Reality-based training platform and a training application based on this concept will be described.

7. Setup and Implementation of the Training Platform

In this chapter the setup and implementation of a multimodal Augmented Reality-based training system based on the concept described in Chapter 5.4 will be described. The description addresses the training platform as well as the implementation of the specified accelerators (see Chapter 5.2) and the training application that has been used for evaluating the system.

7.1. Setup of the Training platform

The hardware setup of the training platform consists of four major components (see Figure 7.1):

- Tablet PC
- Webcam mounted to the tablet PC
- Haptic (vibrotactile) bracelet
- Customized movable mount for the tablet PC

The training application is running on a tablet PC driven by a 1,3 GHz Intel Core2 Duo processor (SU7300). The tablet PC comes with a 4 GB memory module and a graphics card with 512 MB memory. It is equipped with a 12.1"diagonal LED touchscreen with a resolution of 1280×800. The touchscreen allows for finger and pen input. Thus, the user can easily interaction with the training application using his fingers and without the need of additional input devices. Since the tablet PC only provides a front camera, an additional webcam is attached oriented in "the user's view direction" (see Figure 7.2). Thus, the tablet PC can be used as see-through device. The webcam is a low-cost USB webcam with resolution of 2 MP. However, for the developed training application the use of a resolution of 640×480 is sufficient. This configuration has been selected due to the advantages of using tablet PCs as see-through device in Augmented Reality environments described in Chapter 4.2.

The tablet PC can be put on a movable mount that is fixed at a table. The mount has a custom-built holding plate to which the tablet PC can be attached. The tablet PC is secured on the holding plate with hook-and-loop tape, so enabling that it can be easily attached and removed to/from the mount.

Furthermore, the trainee is equipped with the haptic bracelet described in Chapter 6.2.3 for receiving vibrotactile information. The user is wearing the bracelet at his primary forearm, that is right for right-handers and left for left-handers, close to the wrist. This way, the bracelet does not constrain the user in performing operations.

Figure 7.1.: Setup of the training platform: tablet PC equipped with a webcam, movable mount, and haptic bracelet.

7.2. Implementation of the Multimodal Training Platform

As mentioned before, the multimodal training platform must handle the capturing, the processing of the captured data, the application logic such as the selection of instruction data and the generation of adequate feedback, and the presentation of information and feedback to the user. Furthermore, it must allow for accessing the data repository and provide interfaces to connect to the physical components (e.g. cameras or display devices). In the following it is shown how the platform has been realized by adopting the idea of the X3D standard.

7.2.1. Exploiting the X3D Standard

An important factor for the sustainability and portability of the training platform is the use of standard interfaces. Unfortunately, for Mixed Reality there is no standard at all (see [PR06]). Even there is no standard for low-level device-input/output and for constructs similar to the WIMP interface. The same applies for 3D user interfaces. A common approach to handle the scene-management in Mixed Reality frameworks is the utilization of scene-graph concepts such as OpenSG [RVB02], OpenSceneGraph [Ope03], or Performer [RH94]. However, those libraries use a proprietary scene and file structure. In contrast, in the Web3D community the X3D ISO standard [Web08a] has been established. This

Figure 7.2.: The tablet PC is equipped with an additional webcam to act as see-through device.

standard defines features that appear to be beneficial for Mixed Reality environments in general and Augmented Reality in particular:

Scene-Graph X3D provides a directed acyclic graph. This data structure is very commonly used for MR-applications to organize the data and the spatial and logical relations. Also all usual node types to describe transform-, geometry-, and material-objects including shader are included in the standard.

Behavior-Graph X3D allows for modeling the event flow and the application behavior by connecting input and output slots of nodes using *routes*. In contrast to other systems that define different node types for different edge types (e.g. event-flow), the X3D-routes link nodes which are also part of the scene-graph.

Device-independency The X3D standard provides support for intersection-rays and includes device-independent high-level sensors for interactive scene elements. These definitions are in particular useful in combination with low-level sensor data.

Portability X3D specifies a powerful scripting-interface for combining the scene- and behavior-graph. This scripting-interface provides a very flexible runtime environment. For the development of X3D application, nodes are instantiated, scripted, and wired. The resulting networks of nodes are stored in X3D-supported files which are portable between different run-time environments.

Multi-encoding In contrast to the precedent VRML ISO standard that only supported a proprietary encoding and file syntax based on the Inventor toolkit [SC92], the X3D standard includes three different encodings: a VRML-compatible classic encoding, an XML-based encoding, and a binary-compressed encoding.

Due to this richness of the X3D ISO standard, it has been used as basis to design a multimodal training platform that provides a clear separation of system and training application. For the usability of the platform this separation between platform and application is important, since the platform development (e.g. adding new device interfaces) should be completely independent of application development (e.g. implementation of accelerators) or rather from the application content and the training protocol. For this reason, the software platform is realized using low-level programming APIs while application development integrates content, specifies different training protocols, and the describes the application logic using high-level scripting languages. That way, applications can be modified and new applications can be realized by simple scripting operations also from people not exercised in programming (e.g. scientists in psychology of perception). Furthermore, changes in the training protocol and in the specification of the accelerators can be easily realized. This clear separation of platform and applications fosters the sustainable usability of the developed technologies and supports the transfer of the technologies to different training scenarios.

The implementation of the training platform is based on the system framework, now called *instantreality* [Fra08], that adopts the idea of X3D ([BDR04], [BDJW07]). The framework is a VR/AR middleware which utilizes OpenSG [RVB02] for visual rendering and a super-set of X3D nodes as application description language. It provides network-interfaces to incorporate application data at runtime, e.g. from simulator packages or external devices. Furthermore, the framework extends the X3D standard in order to fulfill the requirements of advanced Augmented Reality environments [JFDB07]. Furthermore, it supports standard spatial audio rendering APIs, such as DirectSound and OpenAL.

7.2.2. Application and Data Handling

For the organization of the data and the spatial and logical relations a scene-graph data structure is used. The system framework provides all usual node types to describe transform-, geometry- and material-objects. Beyond that, the implemented application model adopts the X3D concept and extends the traditional scene-graph concept of a node. A node is no longer just a static graph-element but defines a state, state-changes, and input/output slots for every node. Application development is done by instantiating and wiring nodes, which are provided by the system. Thus, development and testing can even be done on regular standard desktop computers and applications can be created by simple scripting operations.

The nodes containing the data (e.g 3D models) communicate with each other via *routes*, which can connect output slots of a node to input slots of other nodes (see Figure 7.3). When an output event is triggered at an output slot, the connected input slot receives an input event. Hence, the event flow of the application and the application behavior, that is the application logic, is realized by connecting input/output-slots of nodes. According to the X3D prototype concept, new node prototypes with behavior scripted in Java or JavaScript can be created. Once defined, prototyped node types may be

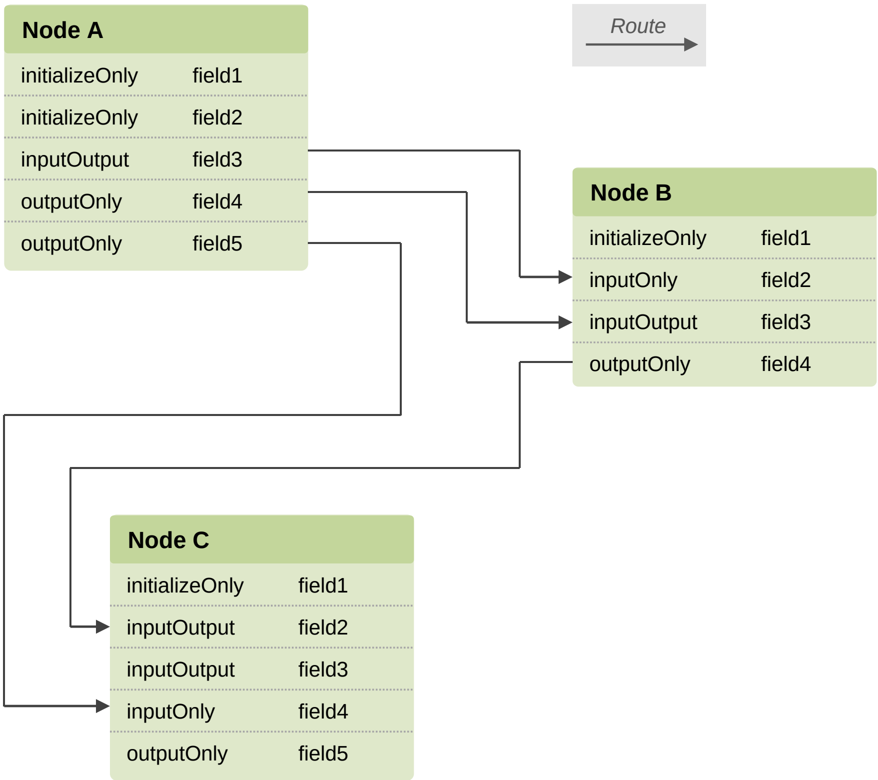


Figure 7.3.: Concept of nodes, fields, and routes: Node communicate via routes. Out-slots of a node can be connected to in-slots of other nodes.

instantiated in the scene-graph exactly like the built-in node types. A prototype node consists of two parts, a declaration (ProtoInterface) and a definition (ProtoBody). The prototype declaration describes fields and field access types for the new node type. This includes the types, names, and default values for the prototype's fields. The definition of a prototype node consists of one or more nodes, which can be nested, and routes connecting these nodes. The first node type in the prototype definition determines the type of the new node. That is, it determines how instantiations of the prototype can be used in an X3D scene. By setting the parameters of the prototype declaration an instance of this node prototype is created. Wherever the prototype is instantiated, a deep copy of the prototype definition scene-graph is inserted in the X3D scene. That way, new nodes describing for instance Adaptive Visual Aids or progress bars must be created only once and can be reused.

To control animations and to handle user interaction and raw data of devices (e.g. cameras) or software components (e.g. tracking libraries) the X3D sensor concept is adopted and extended. The implemented sensors can be distinguished into two categories: low-level sensors and high-level sensors. Low-level sensors allow the application to send or receive raw data streams without imposing any interpretation of these streams by the X3D browser. The application is responsible for handling those data in a useful and meaningful way. For example, if the application receives a video image and the corresponding 3D camera pose, both image and pose must be considered in order to generate a correct Augmented Reality scene.

High-level sensors are sensors that are built on top of low-level sensors. They provide a more abstract interface to devices. Using these sensors the application receives information about the user interaction and can control animations. Thus, the user can easily manipulate the scene. For example, he can transform virtual objects or initiate actions.

Within the application external (native) libraries can be accessed via a *Script* node using the Java Native Interface (JNI). The JNI is a programming framework that allows Java to call methods of platform specific (native) libraries written in other languages (e.g. C, C++). It is a powerful feature of the Java platform that allows Java code running in a Java Virtual Machine (JVM) to call methods of native applications and libraries, such as Windows-DLLs or Linux shared libraries. In the opposite direction the JNI also enables Java code to be called by those applications and libraries which may also be written in other programming languages such as C, C++ or assembly.

The implementation of the *Script* node enables also to communicate with HTTP servers, such as databases communicating over the HTTP protocol.

7.2.3. Device IO and Integration of Software Components

The realization of the capturing and the rendering controller is based on a network transparent transport layer. This layer is used to find and distribute data streams on different systems. Not only simple 3D or 6D streams but even more sophisticated streams such as image streams. Any knowledge about nodes or the scene-graph is not required. The implementation of this transport layer is based on the framework described in [Däh08].

As described in the platform concept in Chapter 5.4, the multimodal capturing controller must be able to handle all preprocessing of the data provided by different capturing systems and devices. The implementation of the transport layer based on the mentioned framework enables to access those devices and to preprocess all provided data. For example, it allows the application for grabbing video frames from a webcam and also to use these frames as input for a tracking module that calculates the camera pose based on video data. Another important feature of the transport layer is that it allows for accessing different machines on the network. Thus, devices can be handled which are not connected to the machine that runs the application.

The layer consists of small software modules (e.g. tracking systems or haptic rendering systems) which are specialized on dedicated tasks. On the one hand those software modules can produce or consume data streams and hence act as device drivers. On the other hand they can assume the function of a filter, which means that they transform incoming data streams. The modules are nodes of a dataflow graph, which communicate with each other via the edges of the graph (also called *routes*; see

Figure 7.4). In the application these prefabricated software modules can be assembled into a data-flow graph.

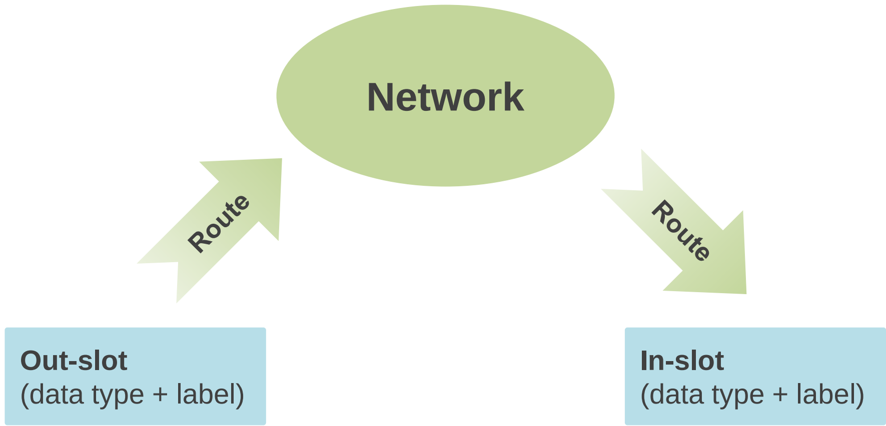


Figure 7.4.: Basic routes concept of the transport layer.

The integration of software components that base on a completely different software architecture, making it difficult to fit them into this data-flow graph concept, is realized using a low-level interface. This interface forms the communication layer of the device management system. It provides channels for a parallelized communication between different software modules. The interface consists of the edges between the nodes of the graph (i.e. it adopts the concept of a data-flow graph). Using this low-level interface events can be injected into the scene-graph. Hence, for the communication between the data-flow graph and a software component whose architecture does not fit in the concept of a data-flow graph, it is not mandatory to write a wrapper node that sticks the component into a node of this graph. Rather those components can use the low-level interface to communicate with the data-flow graph.

Software components that use an architecture similar to the data-flow graph concept can be integrated using the high-level interface, which is build on top of the low-level interface. Using this interface data-flow graphs can be built that consist of nodes that are state machines. Those nodes receive data at their in-slots, change their state depending on the incoming data, and send data on their out-slots. Both low-level and high-level interface are APIs provided by C++ libraries or Java packages.

As already mentioned in the previous section, the visual rendering is implemented based on the open-source scene-graph system OpenSG [RVB02]. Audio and haptic rendering systems can be integrated using the interfaces of the transport layer described above. If no audio rendering system is available, the standard APIs for spatial sounds (e.g. DirectSound, OpenAL) using standard audio hardware are utilized.

7.2.4. Workflow Module

The workflow block consists of different software components which are integrated into the training platform using the interfaces described above. Thus, the workflow components can communicate with other software components (e.g. tracking systems), with devices (e.g. force measuring tool or camera) and with the application. Amongst others, a Dynamic Time Warp (DTW, see e.g. [AV10]) component has been implemented that can be used to measure the similarity between two sequences varying in time or speed (e.g. an expert template and the captured sequence of the trainee). For example, similarities in movement patterns can be detected, even if one time the movement is performed slowly and another time it is performed much faster. The DTW method can also handle accelerations and decelerations in the sequences. If the given sequences are long enough, DTW can also cope with missing information. For the actions, or rather sub-tasks, that should be recognized templates must be created (e.g. by capturing experts performing the actions). The captured performance of the trainee is then aligned with the created templates by warping the sequences non-linearly in the time dimension.

Also a Hidden Markov Model (HMM) component has been integrated into the training platform. Although originally developed for problems in speech recognition [Rab89], HMMs have also proven its potential for the recognition of motion patterns (e.g. [KPKL07], [WKZ08]). To recognize motion patterns a representation is required that represents the movements in a natural way. It has to be taken into account that human performance is inherently stochastic. That is to say, human actions are always characterized by variations, even if the same action is performed twice. For this reason it is not useful to directly model the observation sequences, but rather to model the underlying source for the change in observations. This can be realized using Hidden Markov Models. Since HMMs enable the modeling of spatio-temporal information in a natural way, they establish a good basis for the modeling of motion trajectories. More precisely, HMMs provide a probabilistic framework that can account for dynamically time-varying motion sequences (see Chapter 5.4.3.1).

Beside a low level interface, also a high-level interface for accessing Hidden Markov Models and recognizing motion patterns has been implemented (see also [WKZ08]). Therefor the X3D standard has been extended by additional nodes which allow for easily accessing all parameters of the trained HMMs within the X3D environment:

```
HMM : X3DNode {
  SFString  [in,out]  label  ""
  SFInt32   []        numStates  0
  SFInt32   []        vocabSize  0
  MFInt32   []        vocab  []
  SFInt32   []        delta  -1
  MFDouble  []        initialStateProbs  []
  MFDouble  []        transitionProbs  []
  MFDouble  []        emissionProbs  []
  MFInt32   [in]      sequence
  SFDouble  [out]     forwardProb
}
```

The *label* field defines a label for the HMM. The vocabulary, which contains the amount of possible output symbols, which is equal the size of the codebook, and the symbols itself are given by *vocabSize* and *vocab*. Regarding the possible number of states that can be skipped, a default value of *delta=1* denotes that an arbitrary number of states can be omitted. The *outputOnly* field *forwardProb* returns the forward probability, i.e. the likelihood, of the observation sequence for this HMM applied to the *inputOnly* field *sequence*. The remaining fields contain the re-estimated model parameters.

7.3. Implementation of the Accelerators

In the following, the implementation of the specified accelerators (see Chapter 5.2) within an X3D environment will be described.

7.3.1. Adaptive Visual Aids with Information on Request

For the implementation of this accelerator the visualization concept described in Chapter 5.3.2 has been realized within the X3D environment. Summing up this visualization concept again shortly, Adaptive Visual Aid consists of two major components: a pointer object and a content object. The *pointer* object is spatially aligned (i.e. overlaid on the camera image in the background), while the *content* object is visualized view-aligned.

In order to realize the AVA concept, new X3D prototype nodes have been created. One of those prototype nodes has been implemented to describe the AVA content object, which can contain different types of media, and its behavior. This prototype node provides an interface to set the necessary parameters for defining the media content and ensuring a correct visualization, such as the current window size. The content prototype can handle and display text, images, videos, and 3D content. It consist of a 2D plane as background plane and text objects or textures mapped to 2D objects (i.e. image, video, rendered texture), or both of them, displayed on the top of that plane. The background plane and its content are transformed to the coordinate system of the scene camera in order to enable a view-aligned visualization. If 3D content has to be displayed in the content object, it is rendered into a texture that is visualized on the background plane. The content node can be faded in or out on demand. For toggling the fading state an input slot is provided in the node declaration. The prototype declaration of the new content node looks as follows (X3D classic encoding):

```
PROTO ContentObject [
  field    SFString   title      ""   #title of the content object (e.g. buzzword, step no.)
  field    MFString   addInfo    []   #additional info, if available (e.g. date, author)
  field    SFString   type       "t"  #defines which content should be displayed if more
                                      #than one media field is set: t -> text only,
                                      #i -> image only, v -> video only, r -> 3D content only,
                                      #ti -> text and image, tv -> text and video, etc.
  field    MFString   text       []   #text strings to display (e.g. instructions)
  field    MFString   image      []   #location(s) and file name(s) of images to display
  field    MFString   video      []   #location(s) and file name(s) of videos to display
  field    MFNode     sceneRoot  []   #root node(s) of the 3D content (scene-graph) to display
                                      #(e.g. 3D animation)
  field    SFColor colorBGPlane  0 0 0    #color of the background plane
  field    SFColor colorText     1 1 1
```

```
  field    SFString   dockPosition  "right"  #window side to which the content is docked
  eventIn  SFBool     toggleFadingState      #toggle on TRUE
  eventIn  SFVec2f    setWindowSize
]
```

The pointer object of an Adaptive Visual Aid consists of a predefined 2D or 3D model or animation. No additional behavior description of the pointer model is required. For this reason a new pointer node type has not been defined. In contrast, the Adaptive Visual Aid itself is implemented in a prototype node. The interface of this new node is quite simple. It provides fields for defining the pointer and the content components, and for triggering a toggle of the visibility state of the content object.

```
PROTO AdpativeVisualAid [
  field    SFNode    pointer  NULL  [X3DChildNode]   #root node of the pointer model
  field    SFNode    content  NULL  [ContentObject]  #the content node
  eventIn  SFBool    toggleContentFadingState        #toggle on TRUE
]
```

If a toggle event is received at the *toggleContentFadingState* slot, this event is passed to the *toggle-FadingState* slot of the content node, which then initiates the fading (in or out) of the content plane and all appendant objects. A toggle of the visibility of the content node can also be initiated inside the *AdaptiveVisualAid* node. This happens, if the user clicks on the pointer model or content object (if visible). Those clicks are recognized within the node and a corresponding event is sent to the content object.

In the exemplary training application the pointers of the Adaptive Visual Aids act mainly as area highlight. The content information depends on the current step and can comprise text, images, videos, and 3D animations. The pointer is represented using a pulsing circle or ellipse that highlights the area of interest (see Figure 7.5). Thus, occurring tracking inaccuracies can be compensated. Furthermore, for evaluating the visualization concept of Adaptive Visual Aids the use of a pointer providing less information and a content object providing the detailed information (on request) seems to be useful. That way, it can be examined if the user can understand and cope with this pointer-content metaphor, and if this metaphor is helpful for learning.

7.3.2. Device Display

Two different possibilities for visualizing the *Device Display* have been implemented. The first one is similar to the visualization of the AVA content. It consists of a view-aligned 2D plane and multimedia objects rendered on the top of this plane. The plane can be faded in or out on demand. The objects displayed on the plane provide information about sub-tasks (i.e. steps) belonging to a logical group (i.e. mental group) and the condition of the device before the current step, after the current step, and after the logical group of steps. A text describes the objective of grouped sub-task in a few words. For example, if the grouped steps describe the removal of the actuator cover (e.g. remove screw 1, remove screw 2, etc.), the text "Remove the cover of the actuator" is displayed. An additional object shows either a video of an expert's performance of the sub-tasks, or a 3D animation presenting the sub-tasks

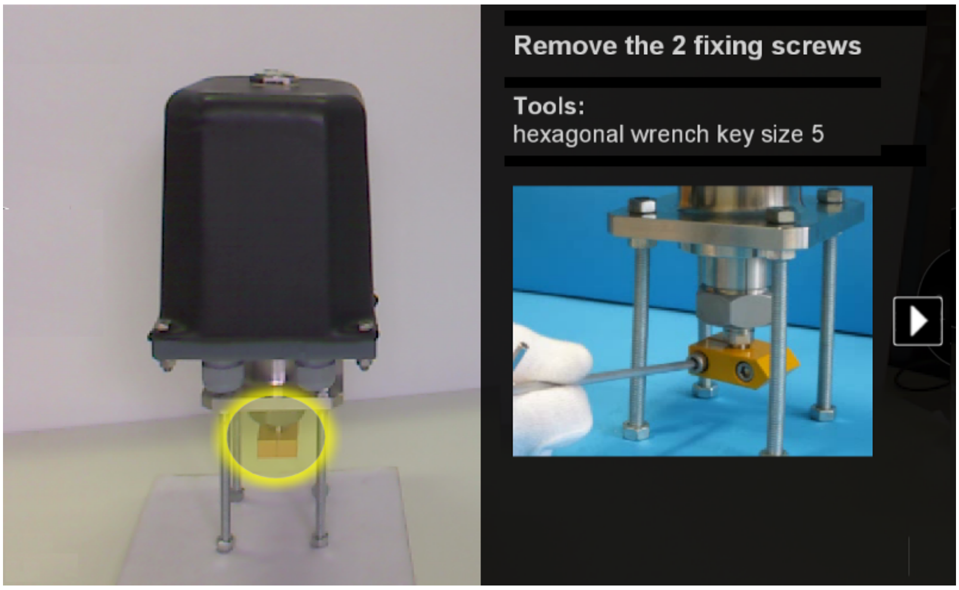
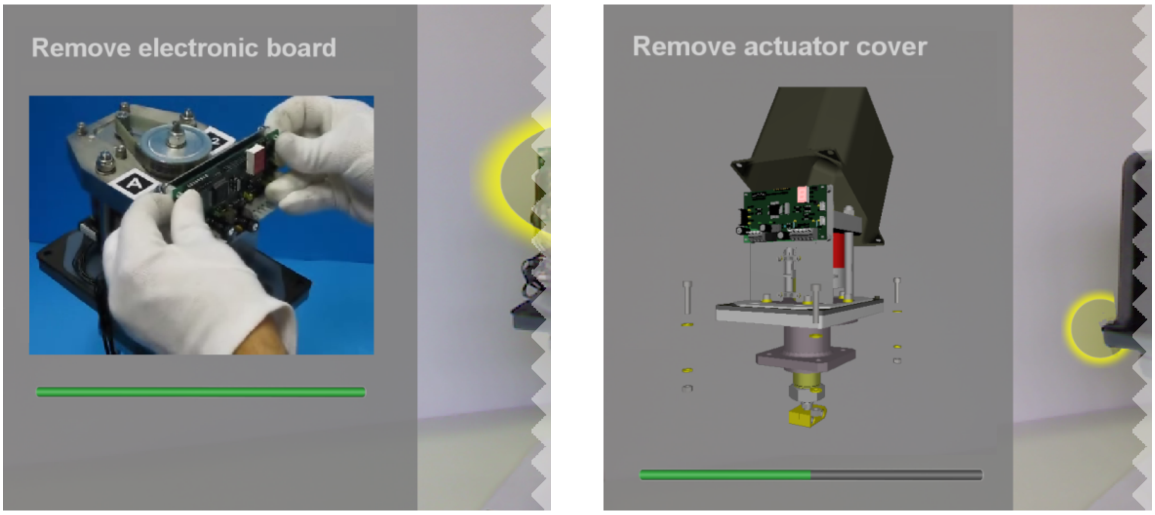


Figure 7.5.: Adaptive Visual Aid in the training application: a pulsing yellow circle (pointer) highlights the area of interest, the detailed instruction is given on the plane (content object).

including the device conditions. Thus, for each mental group the best representation can be chosen. In Figure 7.6 both options are shown. In addition to these objects, also a progress-bar that shows the user's progress inside the mental group is displayed. That way, supplemental information about the structure of the task, or rather of the mental model, is presented, which can further support the user's mental model building process. The implementation of the progress-bar will be described in the next section.

This type of a Device Display is implemented in a prototype node whose interface is very similar to the one of the *ContentObject* prototype:

```
PROTO DeviceDisplay [
  field     SFString   title        "Mental stage:"  #mental stage no. is appended
  field     SFString   text         ""
  field     SFString   video        ""
  field     SFNode     sceneRoot3D  NULL
  field     SFColor    colorBGPlane  0 0 0
  field     SFColor    colorText     1 1 1
  field     SFString   dockPosition  "left"
  eventIn   SFBool     toggleFadingState
  eventIn   SFVec2f    setWindowSize
  eventIn   MFInt32    currentTotalSteps    #set the current step inside the grouped steps
                                            #and the number of steps inside the group;
                                            #used for the progress-bar
]
```



(a) (b)

Figure 7.6.: Device Display presenting the mental groups using (a) a video or (b) an 3D animation.

The fading of the Device Display can be toggled by sending an event to the *toggleFadingState* slot or triggered inside the *DeviceDisplay* node. If the Device Display is visible, a user click on the displayed plane activates a sensor attached to the plane, and as a result the fading out of the plane and its content is initiated.

As mentioned, a second possibility for visualizing the Device Display has been implemented. Similar to the visualization described above, it consists of a transparent view-aligned background object and a 3D animation presenting the grouped steps and the accordant device conditions. Also this Device Display can be faded in and out on demand. However, the background object with the 3D animation only fills a corner of the screen. The most important feature of this type of Device Display is the mapping of the rotation of the tracked camera pose to the 3D animation displayed on the background object. That is, the orientation of the real camera (i.e. the users view) is applied to the scene camera of the 3D animation that is rendered in a texture (see Figure 7.7). In this way, the 3D animation appears in the same orientation than the real device. This facilitates transferring the information into the current situation for the user.

Also for this kind of Device Display an additional prototype node has been implemented. This node contains the transparent background object, a 2D object on which the texture used for rendering the 3D animation is mapped, a progress-bar, and the implementation of the behavior of the node. The 3D animation as well as the orientation of the camera are set in the parameters of the node interface. The handling of the fading mechanism and the progress-bar is the same as described before.

```
PROTO DeviceDisplay_Animation [
```

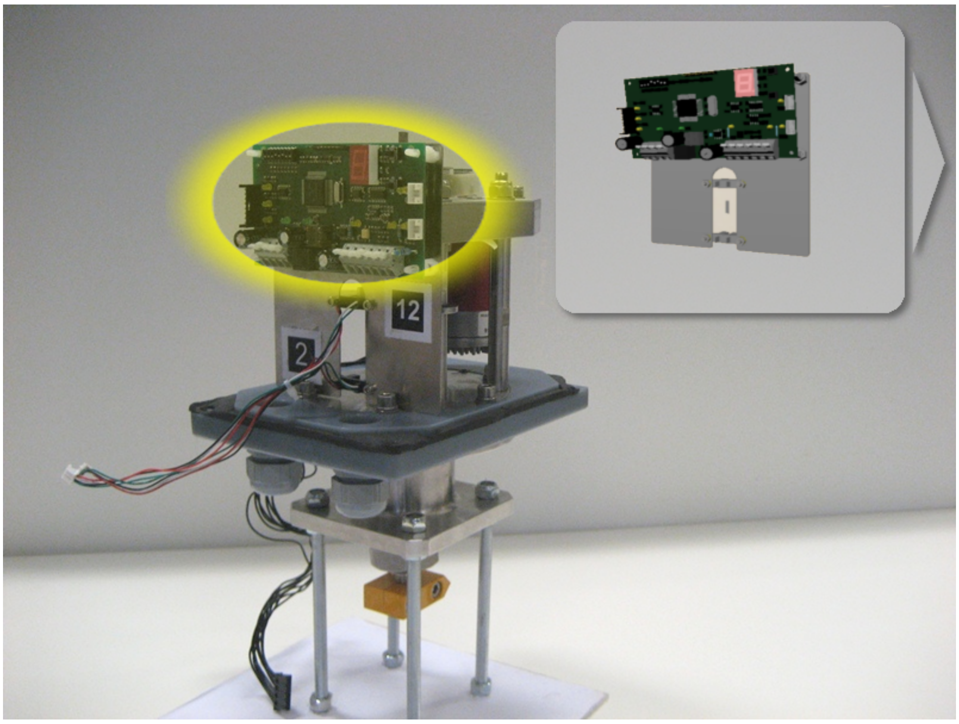


Figure 7.7.: Another visualization of the Device Display: the orientation of the tracked camera is applied to the 3D scene rendered in the Device Display (top right).

```
  field     SFNode       sceneRoot3D    NULL
  field     SFString     dockPosition   "leftTop"   #leftTop, leftBottom,
                                                    #rightTop, rightBottom
  eventIn   SFBool       toggleFadingState
  eventIn   SFVec2f      setWindowSize
  eventIn   MFInt32      currentTotalSteps    #used for the progress-bar
  eventIn   SFRotation   setCameraOrienation  #set camera orientation
]
```

7.3.3. Structure and Progress Information

In order to provide information about the structure and about the user's progress in the task, two different types of progress-bars have been implemented. One is a standard progress-bar that simply shows the user's progress inside a task (see Figure 7.8, top). Such a progress-bar is used in the *DeviceDisplay* node. The other type is an extended progress-bar that additionally provides information about the mental groups and the user's progress inside the mental group (see Figure 7.8, bottom).

Figure 7.8.: Top: a progress-bar showing the user's progress in the task; bottom: an extended progress-bar providing additional information about the mental groups.

An additional *ProgressBar* prototype node has been implemented that can generate both types of progress-bars. Which type is created depends on the parameters which are set in the node declaration when the node is instantiated. The node declaration looks as follows:

```
PROTO ProgressBar [
  field         SFString  dock  ""   #top, bottom
  field         SFVec2f   size  1 0.25
  field         MFInt32   sectionEndpoints []
  exposedField  SFInt32   currentStep 0
  eventIn       SFVec2f   setWindowSize
]
```

The dock position (top or bottom) can be specified by assigning the corresponding value to the *dock* field. The window size, which is a necessary information for docking the progress-bar to the window, is passed to the progress-bar using the *setWindowSize* input field. If no value is assigned, the progress bar is not docked and hence can/must be positioned by the user. In this case, the size of the progress-bar is determined by the value given in the *size* field. The number of steps and the different sections for an extended progress-bar can be determined using the *sectionEndpoints* field. Here, the field value defines the endpoints for the single sections. For example, assigning the field value [5,8,12] defines an extended progress-bar with 12 steps and three sections: from 0 to 5, from 6 to 8, and from 9 to 12. In case the field value contains only one entry, a progress-bar consisting of one section with the number of steps given by the field value is created, which corresponds to a standard progress-bar.

To provide information about the structure of the device (i.e. the actuator), a transparent 3D model of the device can be visualized as overlay in alignment with the real device.

7.3.4. Haptic Hints

For the realization of haptic hints (see Chapter 6.4.1.4), the vibrotactile bracelet described in Chapter 6.2.3 has been integrated into the system framework using the high-level interface of the device manager. That way, it can be easily accessed within the X3D environment by instantiating the corresponding node and connecting the slots:

```
DEF vibroTacSensor IOSensor
{
```

```
  type "VibroTac"
  field    SFString  device "COM3"   #the device which the node should operate
  field    SFString  numModules "1"  #defines the number of modules used
  eventIn  MFFloat   intensity1      #values for the 6 actuators of the bracelet
}
```

The *intensity1* slot receives intensity values to control the intensity of the six actuators inside the bracelet. If less than six values are received, they are mapped to the first actuators in the actuator list of the bracelet. For example, if only three intensity values are given, these values are assigned to the first, second, and third actuator in the bracelet. In case more than six values are received, the first six values are assigned to the actuators. The implementation allows for running multiple bracelets in parallel. This is useful, if two bracelets, one for the left wrist and one for the right wrist, are used in training scenario. In this case, two input slots (intensity1, intensity2) have to be defined for controlling the intensity of the actuators inside the bracelets.

To facilitate controlling the bracelet within the application, an additional node has been prototyped that implements control templates for producing different sensations at the user's arm. The node contains no visible objects. By sending events to the input slots of the node the corresponding methods for calculating the intensity values for the actuators are called inside the node, and the values are assigned to the actuators of the bracelet. An extract of the node interface is shown here:

```
PROTO VibroTac [
  eventIn  SFInt32  knock1        #set number of "knocks" at actuator 1
  ...      ...      ...
  eventIn  SFInt32  knock6
  eventIn  SFInt32  fullKnock     #set number of parallel knocks at all actuators
  eventIn  SFInt32  multiKnock    #set number of parallel knocks at certain actuators
                                  #first value = no. of knocks, rest = IDs of actuators
  eventIn  SFInt32  rotationCW    #set number of clockwise rotation patterns (-1 = inf)
  eventIn  SFInt32  rotationCCW   #set number of counterclockwise rotation patterns
  eventIn  SFVec2f  moveDirection #set direction vector of translation pattern
  field    SFFloat  intensity  0.75      #max intensity for controlling [0..1]
  field    SFInt32  knockDuration  100   #duration of a single knock in ms
  field    SFTime   rotationCycleDuration  1000  #cycle duration in ms
  ...      ...      ...
]
```

By setting the *intensity* field the stimulus intensity can be adjusted. For example, if a low intensity value is set, triggering a knock event would lead to a soft knock stimulus, while a high value would lead to an intensive stimulus. The interface allows for initiating one or more knocks at one particular actuator, simultaneously at multiple actuators, or simultaneously at all actuators. Apart from that, the generation of stimuli producing clockwise or counterclockwise sensations can be triggered. A certain number of rotation cycles can be initiated as well as a continuous rotation. The node declaration also enables to trigger translational stimuli patterns by setting a direction vector to the corresponding input slot (*moveDirection*).

In order to produce different sensations, appropriate vibration intensities must be calculated for the actuators. For generating rotational sensations the actuators are activated according to a sinusoidal wave passing through the actuators. This results in a smoother sensation than it would be if the actuators are activated individually one after another, what facilitates the user's recognition of the motion pattern and of the rotating direction. A schematic diagram about the activation of the actuators in the bracelet is shown in Figure 7.9. Also for calculating the actuator intensities to produce two-dimensional

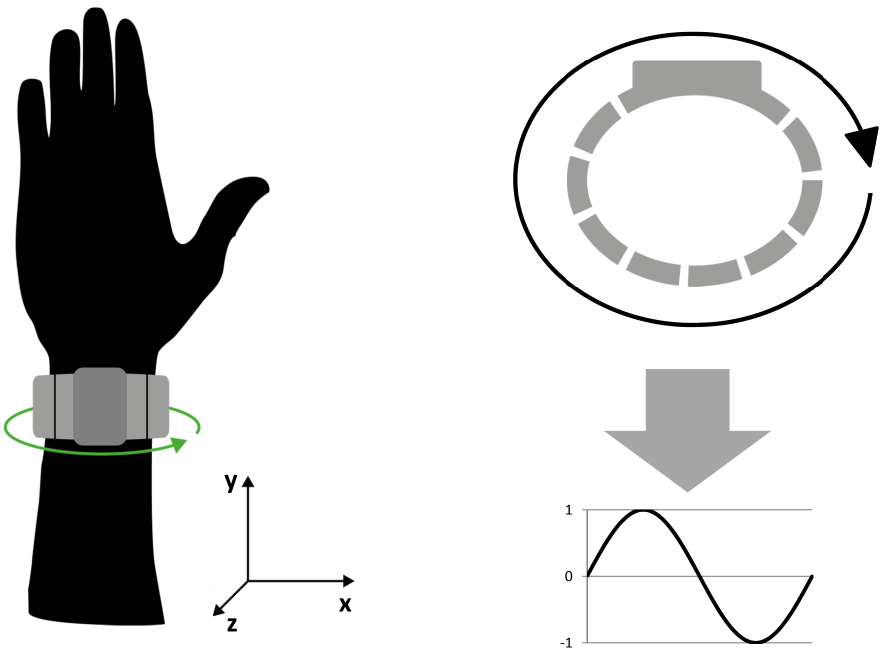


Figure 7.9.: Schematic representation of producing vibrotactile rotational sensations: a sinusoidal wave passing clockwise or counterclockwise through all actuators is used for calculating the actuator intensities.

translational sensations a sine function is used. In this case, two sinusoidal waves starting at the same point pass through the half of the actuators each in opposite direction (see Figure 7.10). The two-dimensional direction vector that is sent to the *moveDirection* slot for initiating the translational stimuli is used for determining start and end point of the sinusoidal waves (i.e. for determining the start and end actuator). The direction vector relates to the coordinate system defined for the bracelet, and hence to the coordinate system of the wrist. An illustration of the generation of translational sensations is given in Figure 7.10.

In the exemplary training application rotational and translational vibration stimuli are used to provide haptic hints about movements to perform. Amongst others, rotational hints are given if screws have to be removed, and translational hints are given if screw nuts have to be removed using a wrench.

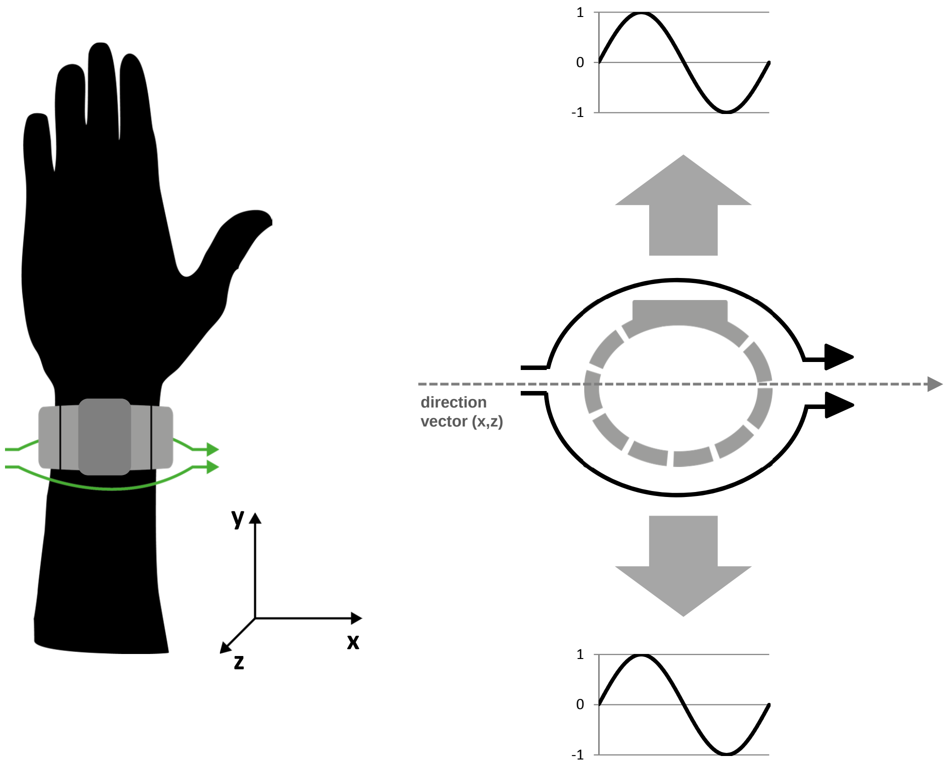


Figure 7.10.: Schematic representation of producing vibrotactile translational sensations (in the xz-plane of the coordinate system of the hand): two sinusoidal wave each passing contrarily through half of the actuators are used for calculating the actuator intensities. The direction (e.g. start end end point) is determined by a given direction vector.

7.4. Implementation of the Training Application

In the following the realization of the training application will be described. Special attention will be given to the application setup, the configuration of the application, the flow of the training procedure and the generation of output providing information about the trainee's actions and performance.

7.4.1. Application Setup

In the application setup all available training data (images, 3D models, mental group definitions, etc.) required for creating the instructions and additional aids is managed. Furthermore, the graphical user interface, the behavior of the application, and the flow of the application are implemented.

In order to manage the instruction data and assign it to the different training phases, a *Step* node has been implemented. For each step of the training task an instance of this node is created, and all instruction data related to this step is assigned to this instance. This data can comprise different pointer models, text descriptions, images, videos, 3D models and more. In is also defined, which data should be visualized in the different training phases. This is done by indexing the different elements and mapping the indices to the training phases. Based on the data assigned to the *Step* node, Adaptive Visual Aids can be generated accordant to the current training phase. The *Step* prototype contains no visible nodes. Rather, it handles the selection of data that is displayed using Adaptive Visual Aids (pointer and content) with respect to a given training phase. Furthermore, the Step node contains information about haptic hints. If no values are assigned to the corresponding fields (*hapticHintCycleDur, transHapticHint, hapticHintIntensity*), no haptic hint is available for this step.

The mapping between data and training phase is done by assigning indices to the given data. For each array of data an index array is defined. The number of entries in the index array corresponds to the number of training phases. That is, the first entry of the index array contains information about the content to be used in the first training phase, and so on. For example, if three different pointer models are given for five different training phases, in the corresponding index array five entries must be defined (i.e. one for each training phase) containing the information which pointer model should be used in this training phase. The data and index array parameters are set in the interface of the *Step* prototype. The interface provides an input slot for setting the current training phase. When this slot is triggered, the data related to the given training phase is assigned to the corresponding output slots. Based on this output, the Adaptive Visual Aid can be generated. An extract of the declaration of the *Step* node is shown here:

```
PROTO Step [
  field    MFNode    pointerModels  []  #pointer models
  field    MFInt32   pointerIndex   []  #pointer indices (e.g [0,1,0,0] for 4 training phases)
  field    MFString  contentTextStrings  []  #instruction text strings
  field    MFString  contentTextIndex    []  #text indices, multiple strings per training
                                             #phase possible (e.g ["0,1,2","1,2","3,4","3"])
  ...      ...       ...
  field    MFString  contentVideos       []  #instruction videos
  field    MFString  contentVideoIndex   []  #image indices, multiple videos per training
                                             #phase possible (e.g ["0,1,2","1,2","3,4","3"])

  field    SFFloat   hapticHintCycleDur   #duration of one rotational cycle (i.e. for one
                                          #rotational hint or one translational hint)
  field    SFVec3f   transHapticHint      #direction vector for translational hint
  field    SFFloat   hapticHintIntensity  #intensity of vibration stimuli

  eventIn  SFInt32   setTrainingPhase     #current training phase
  eventOut SFNode    pointerModel         #returns root node of pointer model
  eventOut MFString  contentTextStrings   #returns text string(s) to display in AVA content
  ...      ...       ...
  eventOut MFString  contentVideos        #returns video(s) to display in AVA content
  ...      ...       ...
]
```

For managing the mental groups another prototype node, called *MentalGroup*, has been implemented. In this node the steps belonging to a mental group and the representation of the grouped steps (i.e.

video or 3D animation) can be defined. Each instance of this node contains the information about one mental group. The node interface is shown here:

```
PROTO MentalGroup [
  field  MFInt32   steps  []          #steps belonging to the mental group (e.g. [3,4,5])
  field  SFString  type   ""          #type of representation if more than one is given
  field  SFString  video  ""          #location of video representing this mental group
  field  SFNode    sceneRoot3D  NULL  #root node of the 3D animation
]
```

The user can interact with the application using graphical user interface (GUI) elements. These elements allow to configure the application, switch between steps, activate and deactivate visual aids, and pause the tracking in order to freeze the current view. Figure 7.11 shows the application with GUI elements during training.

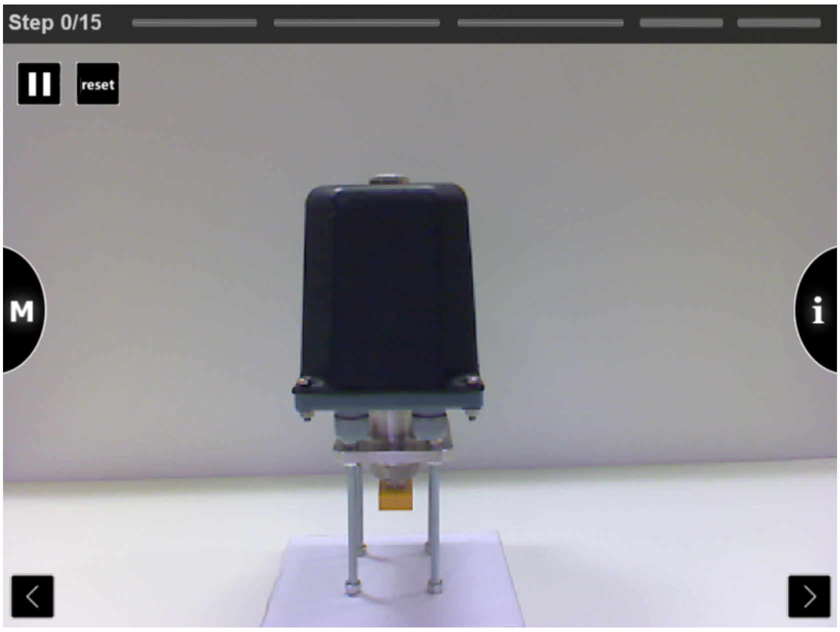


Figure 7.11.: The application with GUI elements: buttons allow for switching steps (<,>), activating visual aids (i) and device display (M), halting the tracking (pause), and reseting the tracking (reset).

The training application covers three main functionalities: the configuration of the training parameters, the execution of the training, and the generation of output in order to analyze and asses the trainee's performance. An overview of the application flow involving these requirements is shown in Figure 7.12. These three main functionalities will be described in the following.

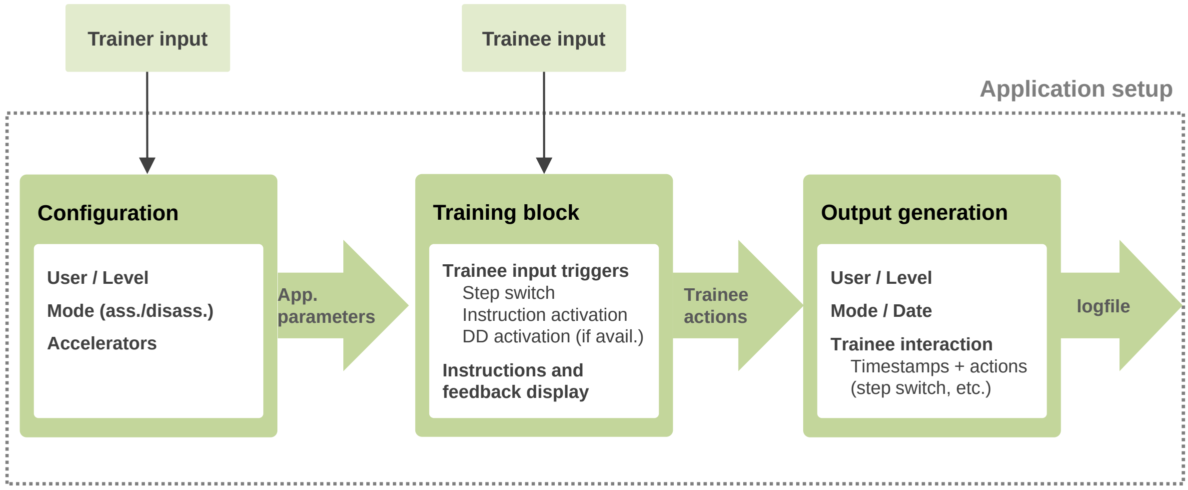


Figure 7.12.: The three main functionality blocks of the training application which are involved in the application flow.

7.4.2. Configuration

The application starts with a configuration menu that allows for configuring the training procedure. In the configuration menu the trainer can define individual training conditions for the trainees. This includes the enabling/disabling of the different accelerators, the setting of the training phase, and the selection of the training mode (assembly/disassembly). These configuration settings can be stored in an user profile. In addition, existing user profiles can be loaded and modified if necessary. Hence, if the trainer is not available on-site during training, he can create a user profile and determine individual training settings for each trainee before the training. When the training start, each trainee can load his user profile and perform the training under the conditions determined by the trainer.

For realizing such interactive configuration dialogs 2D graphical user interfaces (2D GUIs) have proven its worth. 3D GUI elements often fail to gather support for reasons of simplicity and usability. However, the X3D standard does not provide any standard 2D GUI elements. Therefore, we developed a set of nodes that introduces different kinds of 2D user interfaces to X3D [JWO*10]. The proposed node set allows embedding standard 2D UIs, web-content or full applications into 3D scenes. These nodes are used as textures, so they can be bound directly to geometries. A base node has been introduced that defines interactive behavior in texture space. From this node several exemplary UIs have been derived, such as an texture node for self-defined interfaces via a widget toolkit designer tool (e.g. *QT Designer* [Nok11]). The implementation allows for interacting with these textures using the standard X3D pointing (e.g. mouse click) and keyboard input mechanisms. That way, standard 2D GUI elements, such as radio buttons, check boxes, text boxes, or drop-down menus, can easily be used in the X3D environment.

For the configuration menu of the exemplary training application a widget containing the 2D UI elements has been designed using the *Qt Designer* tool. The resulting widget file is mapped as texture onto a 2D plane within the X3D application and rendered parallel to the viewing plane (i.e. view-aligned). Furthermore, tooltips have been implemented which provide detailed information about the configuration options. Figures 7.13 and 7.14 show the implemented configuration menu inside the training application.

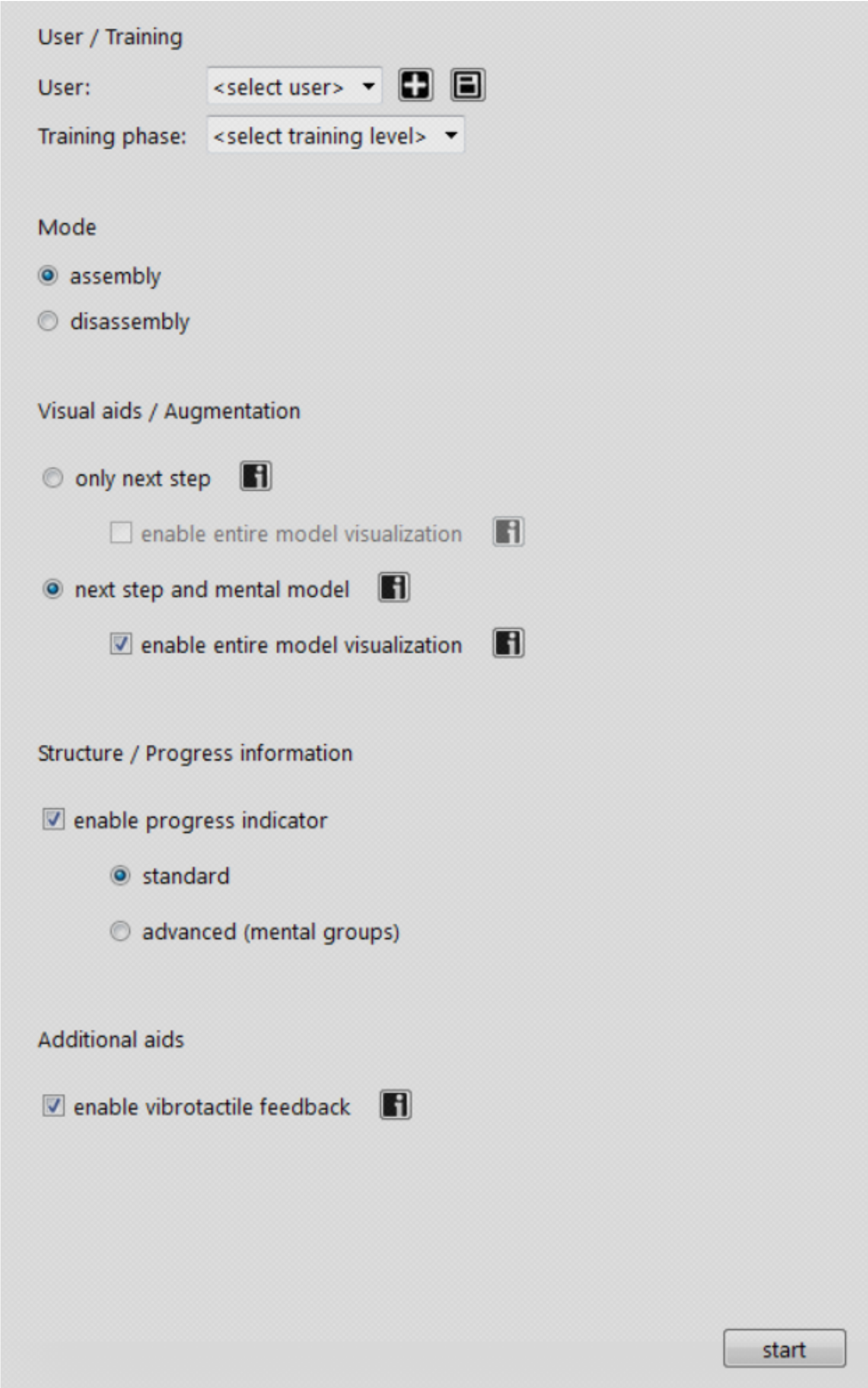


Figure 7.13.: The configuration menu of the training application.

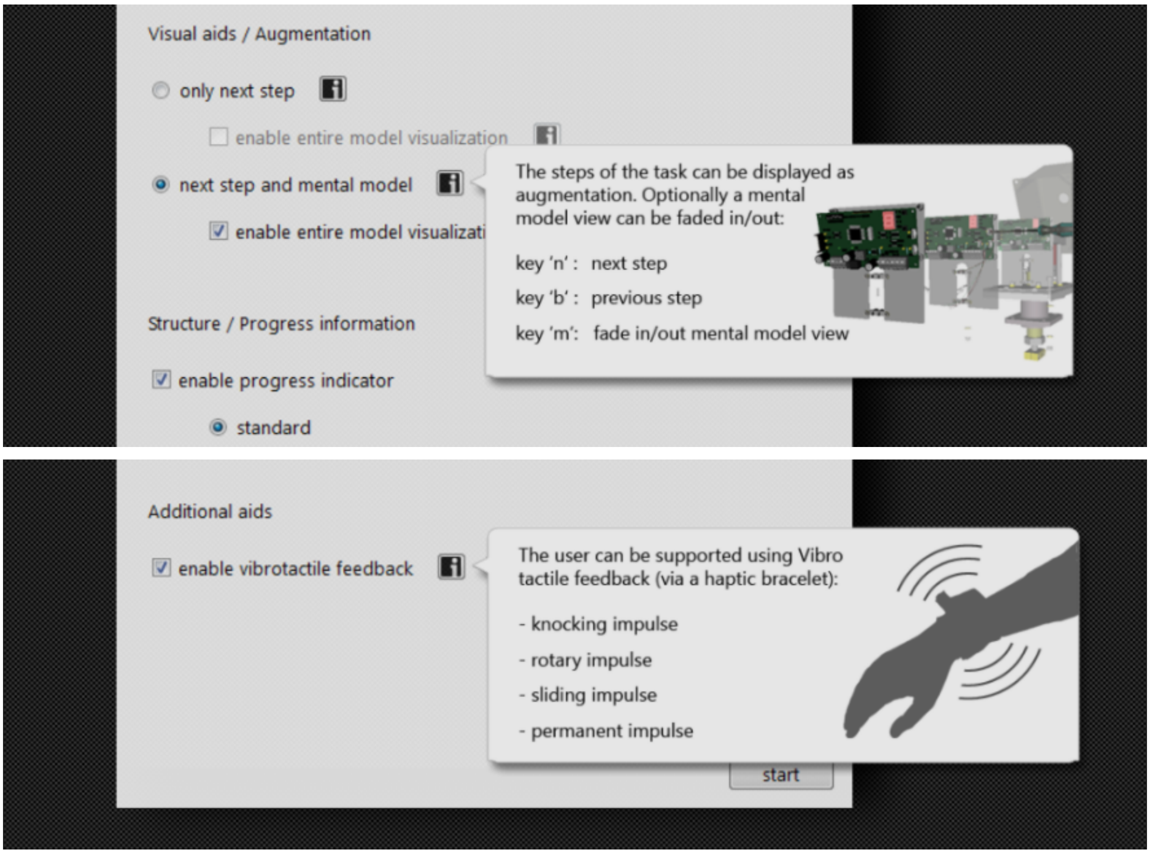


Figure 7.14.: Tooltips provide information about the different configuration options.

7.4.3. Training

When the configuration is finished, at first the presentation of the progress-bar is updated depending on the configured settings. If the presentation of mental model information has been enabled in the configuration menu, an extended progress-bar is displayed. In case no information about mental groups should be provided, a standard progress bar is visualized. If the structure and progress information have been disabled, the progress-bar is hidden.

By activating the corresponding button the trainee can go step-by-step through the training task. He can also switch to the previous step, which is especially important if he realizes that he failed in the execution of the previous step. In this case he needs to go back to this step to correct the mistake. Whenever a new step is activated, an Adaptive Visual Aid is created and, depending on the configured training settings, the progress-bar and the Device Display are updated. Also haptic hints are activated if available.

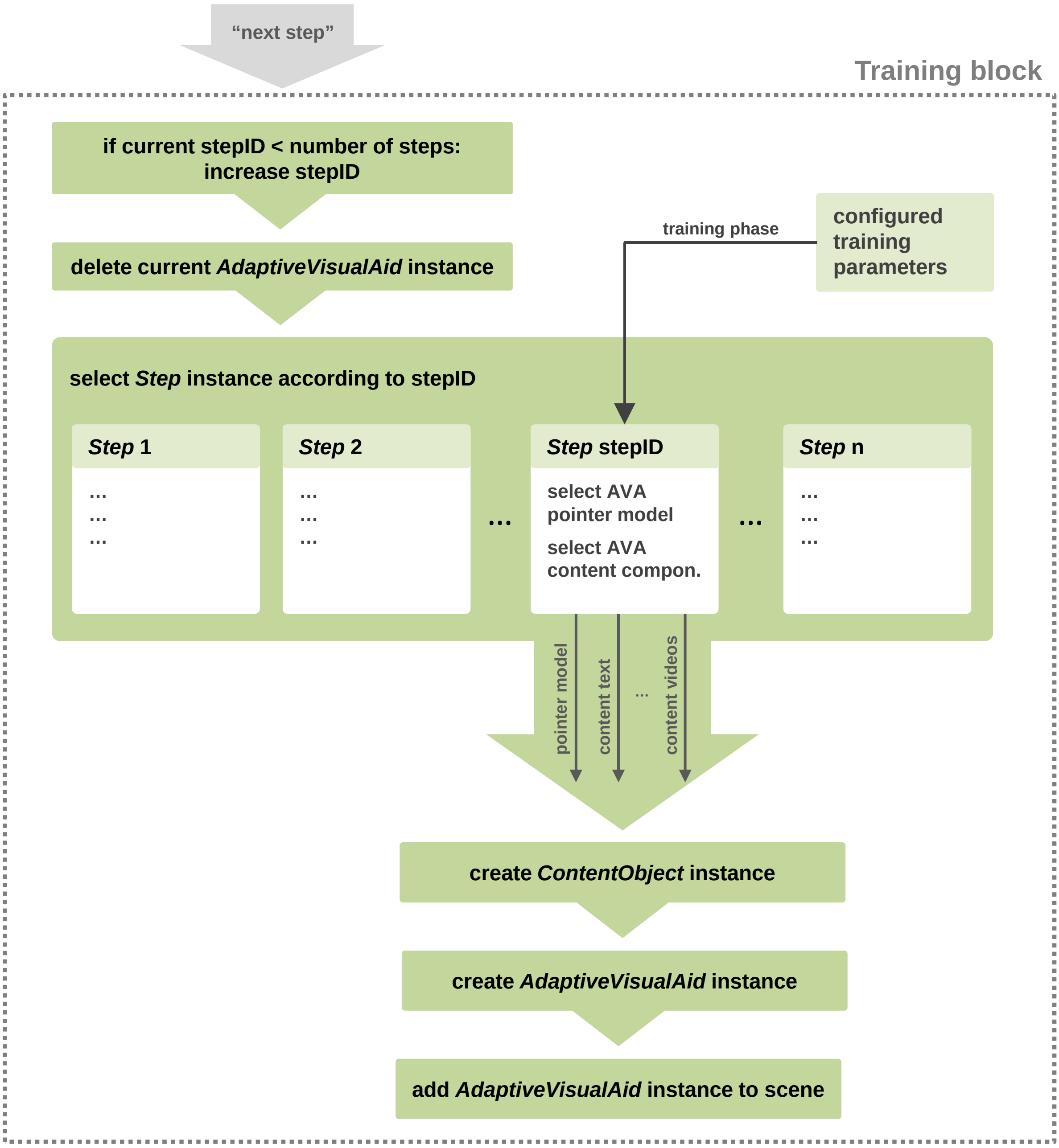


Figure 7.15.: Overview of the process of creating Adaptive Visual Aids and haptic hints when a new step is activated.

The information required for creating an Adaptive Visual Aid is obtained from the corresponding *Step* node instance. By sending the training phase to the *Step* node, the pointer model and the information to be displayed in the AVA content are received. Based on this data the Adaptive Visual Aid can be generated (see also Figure 7.15). As mentioned before, in order to compensate tracking inaccuracies and to evaluate the pointer-content concept, a pointer model is used that acts as area highlight (pulsing circle or ellipse). The content information depends on the current step and can comprise text, images, videos and 3D animations. If the *Step* node instance of the current step contains information about haptic hints according to the current training phase, this information is sent to the *VibroTac* node when the incoming training phase event is triggered, in order to activate the vibration feedback. On each step switch all actuators in the bracelet are deactivated before new values are applied. This ensures the deactivation of the actuators if no vibrotactile feedback is available for this step.

The progress-bar is updated by sending the current step ID to the *ProgressBar* node. The step ID is also sent to the *DeviceDisplay* node for updating its progress-bar. If the current step belongs to another mental group as the previous step, also the main content of the Device Display is updated (i.e. the video or animation). The information that is displayed on the background plane is obtained from the corresponding *MentalGroup* node.

7.4.4. Output Generation

After each training cycle a log file is generated containing the user information (user identification and current training phase), the application settings (configuration of accelerators), and information about the trainee's actions and performance during the task execution. For this purpose, the trainee's interaction with the training application during training is logged. The log information comprises the switching between steps and the activation and deactivation of the content of Adaptive Visual Aids (i.e. the step instruction details), the Device Display (i.e. the mental model visualization), and the visualization of the complete device structure (transparent 3D overlay of the entire device). For each action the timestamp is logged as well. For example, when the trainee activates the content of an Adaptive Visual Aid, the current timestamp, the current step, and the action (AVA content activation) are logged. An extract from an exemplary output file is shown in Figure 7.16. The information in this file is used for analyzing the trainee's performance. This is an important factor for assessing the value of the accelerators and the application. Furthermore, by analyzing the user's actions and performance the next training phase can be determined.

7.5. Summary

To summarize, the multimodal AR-based training platform has been implemented based on the X3D standard, as this standard provides several features that are useful for Augmented Reality environments. Since platform and application are clearly separated, the complete application, including the

User / Date	John Doe	27.04.2011		
Mode / Level	assembly	2		
Configuration	MM:on	EM:on	PI:advanced	VT:on

Time	Diff	Current step	Action	Attribute1	Attribute2
11:57:18	0	1	stepSwitch	next	00:01
11:57:21	2.714	1	stepInformation	on	
11:57:25	4.134	1	stepInformation	off	
11:57:28	3.338	1	mentalModel	on	
11:57:41	12.806	1	mentalModel	off	
11:57:46	5.18	2	stepSwitch	next	01:02
11:57:56	9.333	3	stepSwitch	next	02:03
11:57:59	3.365	3	stepInformation	on	
11:58:10	10.563	3	stepInformation	off	
11:58:11	1.733	4	stepSwitch	next	03:04
11:58:18	6.13	4	stepInformation	on	
11:58:49	31.572	4	stepInformation	off	
11:58:50	0.882	4	mentalModel	on	
11:58:59	8.981	4	mentalModel	off	
11:59:01	1.648	5	stepSwitch	next	04:05
…	…	…	…	…	…

Figure 7.16.: Extract from an exemplary output file that is generated after each training cycle.

accelerators, can be developed within the X3D environment. Thus, it can be easily changed without the need to modify the underlying platform. The implemented training application consists of three major building blocks: configuration, training, and output generation. The configuration component allows for configuring the application (i.e. definition of training phase, accelerators to be applied, etc.), while the training component handles the training process (i.e. step switching, instruction and feedback generation, etc.). Using the output generation component a log-file containing information about the user's actions is produced which can be used for analyzing the user's performance. The evaluation of the implemented training system will be described in the next chapter.

8. Evaluation of the Training Platform

The evaluation of the platform has been conducted at the foods packaging manufacturer *Sidel SpA PARMA*[1] located in Parma (Italy). Amongst others, Sidel develops customized training programs for knowledge and skills transfer in order to improve and maintain the operation of product lines and the performance of equipment and staff. This comprises also the field of technical training.

The specification of evaluation task, evaluation process, and evaluation protocol has been defined in close cooperation with cognitive scientists and employees of Sidel with competence in the field of technical training, as well as the structure and content of all questionnaires and evaluation sheets used for analyzing and assessing the developed training platform.

8.1. General Structure

For the evaluation of the platform, the assembly task of the maintenance scenario described in Chapter 6.4.2 has been used. The actuator to be assembled in the training task is part of a filling station for packaging liquid foods and is produced and maintained at the Sidel location in Parma. Therefore, trainers and trainees, or rather technicians, who need to be trained in working with the actuator are available for the evaluation of the developed training platform. Since the disassembly part of the maintenance scenario is fairly simple for technicians and too easy to learn to draw reliable conclusions about the achieved skill transfer, only the assembly task was selected for evaluating the platform. In order to simplify the evaluation and to allow for deriving reliable statements about the learning effects concerning the procedural aspects of the training task, a fixed order of the assembly sub-tasks has been determined based on the hierarchical analysis of the task described in Chapter 6.4.2 (see Figure 6.5). The order of sub-tasks is defined as follows:

1. **Sub-task D:** Mount the belt roller
2. **Sub-task C:** Mount the two level sensors and the support plate
3. **Sub-task B:** Mount the circuit board
4. **Sub-task A:** Place and fix the actuator cover
5. **Sub-task E:** Mount the clamp to the actuator stem

Also for the single steps of the sub-tasks a fixed order has been determined:

D: Mount the belt roller

[1] Sidel is one of the world's leaders in solutions for packaging liquid foods including water, soft drinks, milk, sensitive beverages, edible oil, beer and alcoholic beverages. `http://www.sidel.com`

1. **D-1:** Join the belt roller to the belt
2. **D-2:** Position the belt roller with the belt
3. **D-3:** Place and tighten the nut that holds the belt roller in place

C: Mount the two level sensors and the support plate

1. **C-1:** Position the upper sensor and tighten the screws fixing it to the support plate
2. **C-2:** Position the lower sensor and tighten the screws fixing it to the support plate
3. **C-3:** Attach the support plate to the structure of the actuator
4. **C-4:** Place and tighten the two screws fixing the support plate to the structure of the actuator

B: Mount the circuit board

1. **B-1:** Attach the circuit board to the support plate by clicking it onto the spacers of the plate
2. **B-2:** Connect the sensor cables to the circuit board
3. **B-3:** Fix the sensor cables together using a cable tie
4. **B-4:** Connect the drive cable to the circuit board

A: Place and fix the actuator cover

1. **A-1:** Place the actuator cover onto the actuator
2. **A-2:** Place and tighten the four screws fixing the cover to the actuator

E: Mount the clamp to the actuator stem

1. **E-1:** Position the clamp
2. **E-2:** Place an tighten the two screws fixing the clamp to the actuator

These steps have been further divided into the single actions that must be performed to accomplish the step (see Appendix C.2.5). As a result, the training task consists of 25 sub-steps that have to be performed one after another.

The evaluation of the platform and hence also of the value of the developed accelerators has been carried out in two stages:

Stage 1: Functionality and usability evaluation by Sidel trainers

Stage 2: Skill transfer evaluation by Sidel trainees

The *functionality and usability evaluation* (stage 1) has been conducted in order to assess the training platform from the "training" point of view. In other words, this evaluation had two basic aims: first, to figure out if the utilized devices are appropriate and easy to use, and secondly, if the amount and type of information provided by the platform during training is adequate and sufficient to understand and learn the training task. Based on the results of the functionality and usability evaluation the platform usability and the specified accelerators have been refined.

The aim of the *skill transfer evaluation* is to assess the training platform as training tool for industrial maintenance activities. For this purpose, the implemented training platform has been compared to traditional training methods by assessing the level of skill transfer achieved through both approaches. In the following, the two evaluation stages and the results will be described in detail.

8.2. Functionality and Usability Evaluation

In this evaluation stage four trainers from Sidel tested the training platform for the selected training task in order to assess the usability of the platform, the pursued training strategies that have been implemented via the accelerators defined in Chapter 6.4.1 (namely, the provision of direct aids, indirect aids and context aids), and the potential of the platform to act as learning tool. The evaluation is not based on performance measures, but on the trainer's subjective assessment of the training platform. The trainers performed one training cycle using the implemented platform and subsequently they had to complete two questionnaires. One questionnaire addresses the usability of the platform and the other one the functionality of the platform. In the usability questionnaire, which is shown in Figure 8.1, the handling of the platform (here called IMA-AR platform for Industrial Maintenance and Assembly – Augmented Reality platform) should be rated by assessing statements concerning the usability of the platform with scores from 1 to 7, where 1 denotes "total disagreement" and 7 denotes "total agreement". Furthermore, an overall grade should be given to the platform (1=very bad, 10=very good).

In the functionality evaluation questionnaire (see Figures 8.2, 8.3, and 8.4) the trainers had to assess the functional properties of the platform. For this purpose, they were asked to give grades for the pursued training strategies implemented through the concrete accelerators and to rate the potential of the platform to act as learning and/or training platform.

"Usability Evaluation" Questionnaire

Instructions: Please assess your experience with the IMA-AR platform. Indicate the extent to which each statement describes you, using a scale in which *1* means you totally disagree and *7* means you totally agree.

1. I was completely concentrated more on the task activities rather than on how to use the system used during the training session.

   1 2 3 4 5 6 7

   Any comment?

2. I felt totally comfortable during the training session.

   1 2 3 4 5 6 7

   Any comment?

3. The interaction with the IMA-AR platform was very easy.

   1 2 3 4 5 6 7

   Any comment?

4. I could ask for information (step information, mental model display) very easy within the application.

   1 2 3 4 5 6 7

   Any comment?

5. I did not have any problem to understand how to use the platform

   1 2 3 4 5 6 7

   Any comment?

6. I think that I am a very proficient user in using the IMA-AR platform after the training session

   1 2 3 4 5 6 7

   Any comment?

7. From the usability point of view: Could you give a grade to the platform in overall?

   1 2 3 4 5 6 7 8 9 10

   Any comment?

8. From the interaction point of view: Which elements of the system did you not like? Which elements of the system were more difficult to understand? Do you suggest any modification or improvement?

Figure 8.1.: The usability evaluation questionnaire used by the trainers to rate the usability of the platform.

"Functionality Evaluation for Training" Questionnaire (1/3)

Instructions: Characterize the functionality from the system from the "training" point of view. Please indicate the extent to which each statement describes you, using a scale in which *1* means you totally disagree and *7* means you totally agree.

1. The information provided by the platform via displayed messages (step description, instructions, etc.) was enough to understand the task.

   1 2 3 4 5 6 7

2. What information would you modify/remove/add it?

3. The visualization of the different operations was enough for learning the task?

   1 2 3 4 5 6 7

4. Which steps would you modify/remove/add?

5. Is there any critical information of the task missing?

6. Could you rate the general visualization utilities: spatial information (i.e. pulsing circles → AVA pointer), step information (i.e. AVA content), captions, etc.?

   1 2 3 4 5 6 7 8 9 10

   Any comment?

7. Could you rate the overview strategy?

   1 2 3 4 5 6 7 8 9 10

   Any comment?

8. Could you rate the direct aids strategy? (spatial information, i.e. pulsing circles (AVA pointer))

   1 2 3 4 5 6 7 8 9 10

   Any comment?

Figure 8.2.: First part of the functionality evaluation questionnaire used by the trainers to assess the functionality of the platform.

"Functionality Evaluation for Training" Questionnaire (2/3)

9. Could you rate the indirect aids strategy? (step information → AVA content)

   1 2 3 4 5 6 7 8 9 10

   Any comment?

10. Could you rate the context aids strategy? (progress bars, Device Display)

    1 2 3 4 5 6 7 8 9 10

    Any comment?

11. Could you rate the haptic hints strategy?

    1 2 3 4 5 6 7 8 9 10

    Any comment?

12. Could you rate the playback/trainer-trainee based strategies?

    1 2 3 4 5 6 7 8 9 10

    Any comment?

13. Could you rate additional features such as: haptic hints, entire transparent model visualisation, configuration possibilities of the demonstrator (i.e. allow Device Display enable haptic hints, use captions always/fade out, etc.)?

    1 2 3 4 5 6 7 8 9 10

    Any comment?

14. From the functionality point of view, could you give a grade to the platform in overall?

    1 2 3 4 5 6 7 8 9 10

    Any comment?

15. What percentage of the task do you consider that you have learnt?

    0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

16. What grade would you give to the IMA-AR platform as learning system?

    1 2 3 4 5 6 7 8 9 10

    Any comment?

Figure 8.3.: Second part of the functionality evaluation questionnaire.

“Functionality Evaluation for Training“ Questionnaire (3/3)

15. What features of the system did you like more?

18. What features of the system did not you like? Do you suggest any modification or improvement?

19. Do you have any additional comment?

20. Do you see any advantage of this training system with respect to the traditional one?

21. Do you see any disadvantage of this training system with respect to the traditional one?

Figure 8.4.: Third part of the functionality evaluation questionnaire.

8.2.1. Results and Analysis

An overview of how the trainer’s rated the usability of the implemented Augmented Reality training platform is presented in the diagrams of Figure 8.5. Figure 8.5 shows the average scores (green blocks) and the range of scores (black vertical line above/below the green blocks) given by the four trainers for the first six statements of the usability evaluation questionnaire. The overall grades for the usability of the platform are presented are presented as well. All average scores for the statements concerning the usability of the platform range between 6 and 7 with a maximum score of 7. These results show that the implemented platform is adequate and easy to use and to understand. This is important to allow the trainees for concentrating on the actual training instead of paying too much attention to the handling of the platform. Furthermore, the trainers stated that the interaction with the platform or rather with the application is very easy. That is, the user can easily initiate the presentation of additional information (e.g. requesting for more step information (AVA content) or for the Device Display).

The average overall grade for the usability of the platform is 7.5 with a minimum grade of 4 and a maximum grade of 9. This shows. Hence, also the overall rating indicates a very good usability of the training platform. The higher grades are in line with the scores given to the usability statements, while

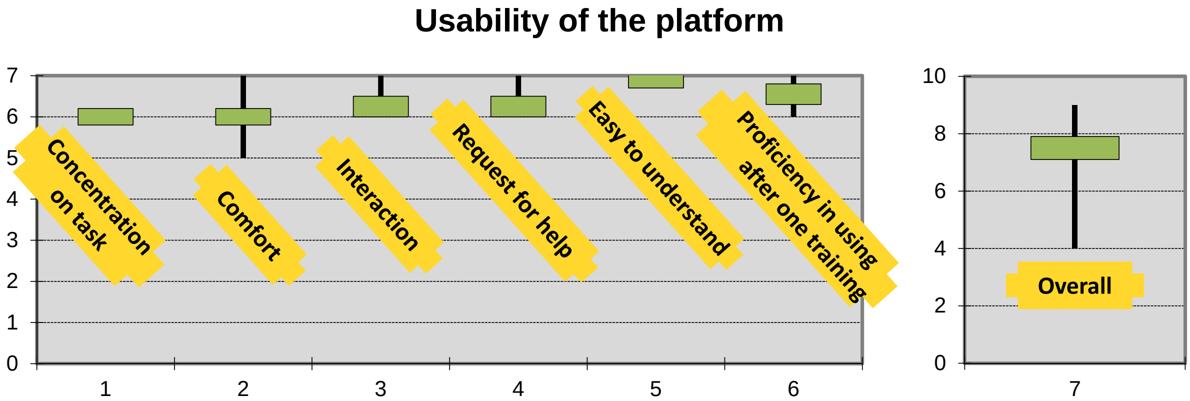


Figure 8.5.: Overview of (left) the ratings for the usability of the training platform provided by the trainers in the usability evaluation questionnaire and (right) the overall grades given to the platform by the trainers.

the minimum score of 4 seems not to fit in this series, since the trainer who gave the overall grade of 4 rated the usability statements with an average score of about 6.2 and a minimum of 5 (where the highest possible score is 7).

In the last point of the usability questionnaire the trainees had the possibility to add comments about the interaction features they liked or disliked and about elements of the platform that were difficult to understand. Here, one trainer noted that the tracking sometimes got lost and could hardly be reinitialized if the marker rectangle that is used for initializing the tracking was partly covered (e.g. by tools that are put back on the table after using). For this reason the user interface of the training application has been expanded by the *pause* button, which allows for pausing the tracking and "freezing" the current view (while the rendering loop is not paused, i.e. 3D animations are still running) once a good perspective of the device for the step is found. This way, the user does not have to pay attention if something is covering the marker when he wants to request information at any time during the performance of a step, as he does not have to reinitialize the tracking. Only when a reinitialization of the tracking is necessary, for example to find a better view on the device for catching the spatial information (i.e. the AVA pointer) in another step of the task. However, this occlusion problem is rather caused by the tracking method that is used to determine the camera pose than by the training platform and application.

Further improvements or modifications of the training platform were not requested. An detailed overview of the usability evaluation is available in the Appendix C.1).

An overview of the results of the functionality evaluation is presented in Figure 8.6. The figure shows the scores for the provided information and the ratings for the implemented training strategies (implemented by accelerators) and the haptic aids, as well as the overall grades for the functionality of the platform and its learning aspects.

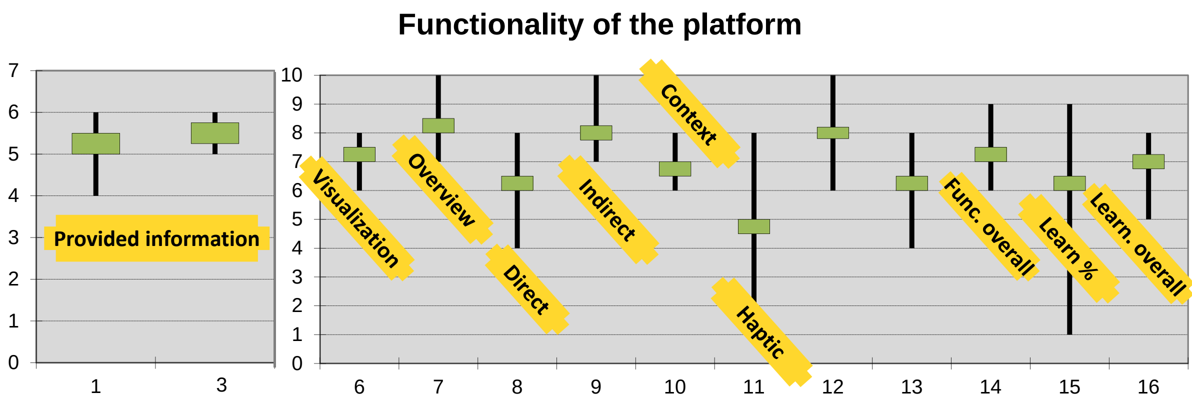


Figure 8.6.: Overview of the ratings for the functionality of the training platform provided by the trainers in the functionality evaluation questionnaire. Left: the grades for amount of provided information; right: the ratings for the training strategies and accelerators and the overall grades for the functionality and learning aspects of the platform.

The average score for the information provided via messages (i.e. step descriptions, instructions, etc.) is 5.25 with a minimum of 4, a maximum of 6, and a highest possible score of 7. The information presented via figurative visualizations of operations (e.g. videos, 3D animations) is rated with an average score of 5.5 with a minimum of 5 and a maximum of 7. These results show, that the amount of provided information was already reasonable, but can be further improved. On this account, the trainers have been interviewed about which messages and figurative visualizations are imprecise or deficient, and accordingly the corresponding messages and visualizations have been refined. Also the possibility for pausing videos in AVA content objects has been added to the platform. Besides, one trainer noted that an overview of the filling process is missing, but as the filling process is not part of the training task, it has been decided not to add information concerning this process into the platform.

The average ratings for the general visualization utilities (e.g. spatial information, step information), for the direct aids strategy (pulsing ellipse, i.e. AVA pointer), for the indirect aids strategy (step information, i.e. AVA content, and haptic hints), and for the context aids strategy (progress bars, Device Display) are between 6.25 and 8 of possible 10. Hence, the trainers see potential in the applied visualization tools and the training strategies implemented via the accelerators. However, one trainer rated the direct aids strategy with a low score of 4. Interviewing him pointed out that he gave this low score due to the tracking problem described above and not because he disliked the direct aids strategy itself (i.e. the idea of the AVA pointer).

The ratings for the provided haptic hints range from 2 to 8 with an average score of 4.75. When consulting the two trainers who gave the lowest scores it became apparent that they see great potential in the idea of providing additional information via haptic hints, but that the actions for which the haptic hints are currently provided (e.g. rotational hints when screws need to be loosened) are very simple and do not require additional hints. They prefer presenting haptic hints for more complex actions.

As no such complicated actions are required in the current assembly scenario, no changes have been made in the application. Also it was not necessary to modify the implemented interface for the haptic bracelet, since it already allows for independently controlling the six actuators inside the bracelet and hence for creating additional vibration patterns. All trainers have been interviewed about the strategy of presenting additional step information using the haptic hints and the haptic bracelet in particular. Here all trainers stated that they assess the idea of providing haptic hints in terms of vibrotactile stimuli as good and valuable for maintenance/assembly tasks, since the information can be provided directly at the location of interest (e.g. at the body part that needs to perform an action). Besides, they consider haptic hints as good for presenting information that is difficult to visualize (i.e. rotating directions). In particular, they like the haptic bracelet, since it is easy to use and does not constrain the user in performing assembly operations.

The scores of overview information that comes along with the provided structure and context information are high (average 8.25). This shows that the trainers consider the provision of information about the structure of the task and about the context of the current step to perform to be a good for understanding and learning the task. Also the trainer-trainee strategy (i.e. configuration, starting, and repeatability of the training) has been assessed as with a very good average score of 8. Moreover, the average rating of the configuration possibilities of the training application and the additional minor features (e.g. transparent 3D model overlay) is considered to be sufficient.

The overall grades for the functional aspects of the implemented training platform and its potential to act as learning system are high (about 7.25 and 7). So the platform could be used for the skill transfer evaluation without having to make any major changes. The trainers assessed the learning rate of the platform after one training cycle with an average of 62.5%, wherby one trainer rated the learning rate with 90%, one with 80&, one with 70%, and one with only 10%. The low value of 10% seems not to be in line with all the other scores given to the platform by the same trainer, such as the scores for the functional aspects of the platform (8) or for the direct aids, indirect aids and context aids strategies (7, 8, 8; see also the detailed functionality evaluation sheet in the Appendix C.1.

Furthermore, the completed questionnaires show that the trainers like the interaction with the platform and the fact that no trainer is required on-site during training. In addition, they noted that the platform is not as error-prone as humans. However, one trainer criticized that the platform does not provide feedback about minor errors. Besides, it has been advised to provide feedback if the tracking is lost. It has also been assumed by one trainer that the platform is good only for simple operations. Consulting him regarding this comment showed that he assumes the platform to be good for tasks consisting of (many) steps which are easy to perform (i.e. which require the performance of simple motor actions). Since most of the maintenance tasks consist of many steps that are themselves easy to accomplish for technicians, the platform is assessed as good for the majority of maintenance tasks.

8.2.2. Summary

To sum up, the subjective functionality and usability evaluation has pointed out that the implemented training platform and the pursued training strategies have been assessed by four trainers as appropriate for the training of maintenance and assembly tasks. According to advises of the trainers, some minor functionalities have been added and several step descriptions have been refined in order to improve the training platform. The updated platform has been used for the skill transfer evaluation, which is described in the next section.

8.3. Skill Transfer Evaluation

The aim of this evaluation stage is to analyze and assess the value of the developed training platform as tool for the training of industrial maintenance and assembly activities. For this purpose, the platform is compared to the traditional training methods that are currently applied in training sessions teaching the selected assembly task. The skill transfer evaluation is conducted with technicians from Sidel with at least two years of practical experience in the field of maintenance and assembly. These technicians also indicated that they are familiarized with the use of personal computers and hi-tech products in general. The participants did not receive any payment for their participation, but the best performers were published at Sidel. As the technicians are part of Sidel's EOL (End-Of-Line) division, only few of them have seen the actuator before. No one of the technicians worked with the actuator before the evaluation.

The evaluation is based on subjective assessments and the analysis of performance measurements taken during training and during performances of the trained assembly task. Training process and performance measures have been selected according to the training protocol described in Chapter 6.4.2.

8.3.1. Evaluation Protocol

The skill transfer evaluation has been conducted with twenty technicians from Sidel, one female and nineteen males. This reflects the overall gender distribution of technicians at Sidel. The participants were divided into two groups each consisting of ten technicians: *AR group* and *control group*. The participants of the *AR group* used the developed training platform during training, while the control group performed the traditional training of the assembly task. Before training, the current skill level of the participants has been tested through a capability test based on a simple assembly task performed with a specific set of modular components for mechanical engineering (*items*, see also Figure 8.8).

After the training, or rather after the training session of each participant, the evaluator filled in a score sheet with the relevant performance measures (completion times, number of errors, number of requested aids, etc.) and comments of the trainees that can be useful to analyze and assess the results. The relevant performance measures have been delivered by the platform (log file) and by the evaluator (through observations and video analysis).

8.3.1.1. Design and Participants

The evaluation experiment follows a between-participants design. That is, each participant has been assigned to only one group. For the allocation of the groups, the participants had to complete a demographic questionnaire (see Appendix C.2.3) concerning their experience, skill level, and familiarity with computer-based (training) systems. This information has been used to distribute the users along the two groups in an homogeneous way. Accordingly, for each member of the control group there is a member with the same (or very similar) skill level in the AR group.

The resulting groups had the following properties: The average time of experience in the field of maintenance and assembly is about 3.8 years for the participants of the AR group and 4 years for the members of the control group, while each group has one participant with more than 10 years of experience. The mean age of the AR group members is about 37.2 years, where the youngest technicians is 27 years old and the oldest one is 43 years old. This distribution is similar to the one of the control group, where the average is 38.8 years with a minimum age of 28 years and a maximum age of 50 years.

The overview of the results of the demographic questionnaires are presented in Figure 8.7. The detailed demographic evaluation sheet can be found in Appendix C.2.3.

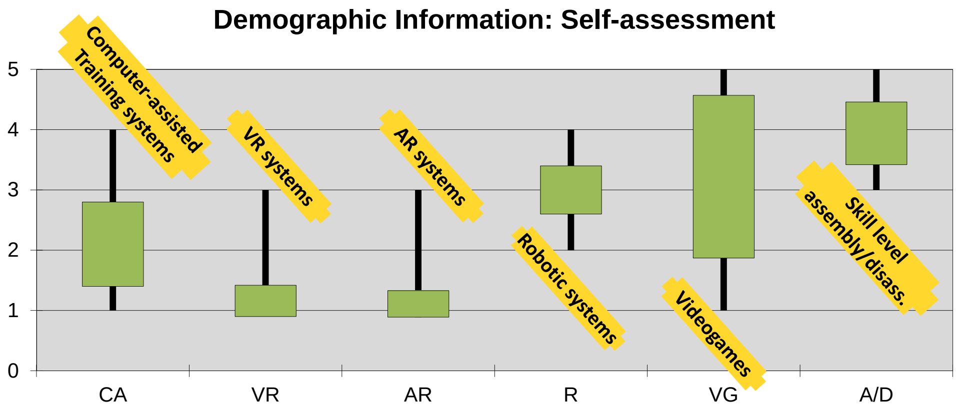


Figure 8.7.: Overview of the results of the demographic questionnaires completed by the participants before training.

8.3.1.2. Material and Setup

As mentioned, the evaluation has been conducted in the Training Center at Sidel. For the skill transfer evaluation the following material has been required:

- The item set for the capability test
- Traditional training material for the training of the control group (i.e. original instruction videos)

- A book of photographs to be used as aids during the evaluation session
- A video camera to record the trainees' actions/performance of the assembly task
- A copy of the evaluation protocol document for the evaluator (see Appendix C.2.1). This document contains the setup and description of the evaluation process, as well as instructions for the evaluator which are important for conducting the evaluation with consistent conditions for all participants (e.g. speech instructions to introduce the evaluation and the training platform).
- Handouts for the platform familiarization session
- Copies of all questionnaires for the trainees
- Copies of a previously prepared evaluation sheets used by the evaluator to record comments and measures during the evaluation (see Appendix C.2.5)
- The physical device to assembly (i.e. the actuator)
- The tools required to perform the training/evaluation task. The tools have to be put on a table in front of the trainee in a predefined order to assure consistent conditions for all participants.

The actuator and the tools have not been visible for the participants until the need to use them.

8.3.1.3. Capability Test

As already mentioned, the participants initial knowledge and skill level is tested by means of a capability test using a specific set of modular components for mechanical engineering, known as *items*. In this test the technicians have been asked to complete a predefined simple assembly task using the *items*. The utilized set of *items* and the structure to be assembled during the capability test are shown in Figure 8.8.

(a) (b)

Figure 8.8.: (a) The set of *items* used for the capability test. (b) The *item* structure to be assembled during the capability test.

For assessing the participant's performance, the time needed to complete the whole assembly, starting from the situation shown in Figure 8.8 (a) and finishing when the structure shown in Figure 8.8 (b) is reached, and the number of solved and non-solved errors he made during assembly have been recorded.

8.3.1.4. Procedure

The evaluation has been undertaken in two parts: one in the morning and one in the afternoon of the same day. Hence, each participant performed two sessions: a training session in the morning and an evaluation session in the afternoon. As mentioned in Chapter 6.4.2, for achieving reliable results concerning the acquisition of procedural knowledge, it is necessary to keep some time between the training and the next time the trained skill/knowledge must be accessed.

The *training session* conducted in the morning was structured as follows:

1. **Capability test:** First, the participant performed the capability test.
2. **Introduction of the evaluation:** The purpose of the evaluation, the evaluation process and the training/evaluation task were introduced to the participant by the evaluator.
3. **Introduction of the training method:** The evaluator introduced the training method used by the participant to train the assembly task. That is, for participants of the AR group the AR-based training platform was introduced; for participants of the control group that traditional training method was presented.
4. **Familiarization session:** For members of the AR group a familiarization session was provided in which the participant learned how to handle the devices, how to use the training platform, and how to interact with the application (see also Chapter 6.4.2). Thus, the trainee could concentrate on performing the assembly task during training instead of paying a lot of attention to the handling of the training platform. For the familiarization session a short document explaining the basics of the training platform was provided (see Appendix C.2.2). Also the evaluator could be consulted for receiving further information. The familiarization session was finished when the participant decided to be sufficiently familiar with the platform to use it for training an unknown assembly task. The duration of the familiarization session was recorded in the evaluation sheet. For the members of the control group no familiarization was required.
5. **Task training:** The participant had to perform one training cycle using the assigned training method. That is, members of the AR grouped performed the assembly task one time using the AR-based training platform, while the members of the control group trained the task using training/instruction videos provided by Sidel. The participant's performance and actions (e.g. requests for step information or Device Display) were recorded via the log file of the training platform and a video camera for further analysis. Also for control group participants the physical device (i.e. the actuator) was provided, even though they were free to decide whether to use it during training or not. In other words, they had the possibility to physically perform the assembly task during training, but it was not mandatory to do it. One training cycle of the control group

was completed when the participant worked through the provided training material one time from start to finish. The time needed for one training cycle was recorded in the evaluation sheet.

The evaluation session which takes place in the afternoon consist of two major parts:

1. **Evaluation test:** The evaluation test consist of the physical execution of the training task without the use of any training tool. The participant was told to perform the trained assembly task without any help (i.e. without using the AR-based training platform or any training material). In case the participant was not able to accomplish a step without help, he could use a book of photographs providing figurative information about the single steps of the assembly task. A video camera recorded the participant during the execution of the assembly task. By analyzing the video after the evaluation session, the evaluator recorded the performance time, the number of solved and non-solved errors, and the trainee's comments that can be useful for assessing the training platform in the evaluation sheet. In addition, each consultation of the book of photographs was recorded as "request for aid". The maximum time allowed for accomplishing the evaluation is 20 minutes.
2. **Completion of exploitation questionnaire:** If the participant was member of the AR group, he had to complete the exploitation questionnaire for the subjective assessment of the training platform (see Figures 8.9, and 8.10). Members of the control group did not have to fill in any questionnaire.

Platform's help to learn the task					
1. Rate the degree to which you completed the required task after the training on the platform *(from 1: not at all; to 5: wholly completed)*	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
2. Do you think the training on the platform quickly enhanced your skill level in the trained task? *(Yes/No)*	Yes ☐		No ☐		
3. Overall, I am satisfied with how I can perform the task trained by the platform *(from 1: strongly disagree to 5: strongly agree)*	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
4. I would recommend the platform for training the task I experienced *(from 1: strongly disagree to 5: strongly agree)*	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
5. I think the platform is a valuable training means *(from 1: strongly disagree to 5: strongly agree)*	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
Any comment?					

Figure 8.9.: First part of the exploitation questionnaire.

Ease of use of the Platform					
1. It took a long practice-time before I understood how the platform works *(from 1: strongly agree to 5: strongly disagree)*	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
2. I was completely concentrated more on the task activities rather than on how to use the platform used during the training session *(from 1: strongly disagree to 5: strongly agree)*	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
3. Rate the difficult you experienced in carrying out the assigned task on the platform *(from 1: very difficult to 5: easy)*	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
4. How efficiently did you complete the required task on the training platform following the indications provided? *(from 1: inefficiently to 5: very efficiently)*	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
5. Rate the level of clarity of the instructions provided during the training on the platform? *(from 1: very unclear to 5: totally clear)*	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
6. Whenever you make a mistake using the platform,					
a. to which extent could you recover easily? *(from 1: difficult to 5:very easily)*	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
b. to which extent could you recover quickly? *(from 1: very slowly to 5: very quickly)*	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
7. Overall, I felt the experience with the platform comfortable *(from 1: strongly disagree to 5: strongly agree)*	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
8. Overall, the platform is easy-to-use. *(from 1: strongly disagree to 5: strongly agree)*	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
Do you suggest any modification or improvement?					

Figure 8.10.: Second part of the exploitation questionnaire.

8.3.1.5. Performance Measures and Scoring

For measuring the participant's performance the measures proposed in the exemplary training protocol in Chapter 6.4.2 have been recorded during the AR-based training and the participant's execution of the task in the evaluation session. These measures are:

- Task completion time
- Requests for aid
- Performance errors

Also the step completion times are measured and used for analyzing which steps caused problems. The performance errors are are divided into three different categories:

- Non-solved errors
- Errors solved with aids
- Errors solved without aids

Furthermore, four different types of error have been defined. All errors made during the performance of the task are assigned to one of the following groups of error types in order to analyze the most common types of error:

- "Forget a piece": e.g. forget the washers
- "Exchange positions of pieces of the same sub-step": e.g. exchange the positions of washers and lock washers when fixing the cover
- "Exchange positions of pieces of the same sub-step": e.g. place screws for fixing the cover in the clamp
- "Wrong positioning of components": e.g. wrongly oriented sensors

Since the aim of the training task is to assemble the actuator and getting a functional device (i.e. all components are well placed and fixed), the number of non-solved errors is considered as the most important performance measure. Also the number of requests for aid are important, as the task shall be performed with the minimum number of aids (e.g. consulted photographs). Fast performances are not irrelevant, but should be considered less important.

Based on these criteria an additional scoring formula has been defined:

$$Score = (20 - TCT) - 4 \times (no.\ of\ requests\ for\ aid) - 2 \times (no.\ of\ unsolved\ errors) \tag{8.1}$$

Here, *TCT* is the task completion time in minutes. The *requests for aid* denote the number photographs consulted during the execution of the task in the evaluation session. The task completion time and the weighted numbers of aid requests and non-solved errors are subtracted from the 20, since 20 minutes is the time limit for accomplishing the assembly task in the evaluation session. The requests for aid (i.e. the use of photographs) have a double penalization than the non-solved errors. The main reason for doing so is that one consultation of the photographs can help for more than the current step if there are similar steps, and hence can avoid performing multiple errors (e.g. consulting the photograph for the mounting of one level sensor helps also for mounting the second level sensor). Furthermore, in the ideal case the task is performed without any help, and hence the participants should be motivated not to request for aid (i.e. not to use the book of photographs). Using this formula the most important performance measures are set in correlation. Thus, an additional score is determined, that can be used for assessing the participant's performance—the higher the score, the better the skill level.

The problem with this formula is that the participants can calculate the costs of a non-solved error (in this case 2 minutes) and based on this they can decide not to correct an error to spend less time. For this reason, formula 8.1 has been reformulated in another form, in which it is presented to the participants:

$$\begin{aligned} Score &= (20 - TCT) - 4 \times RFA - 2 \times NE \\ Score &= TimeScore - 2 \times RFA - 1 \times NE, \text{ with} \\ TimeScore &= (20 - TCT) - 2 \times RFA - 1 \times NE \end{aligned} \quad (8.2)$$

Here, *TCT* denotes the task completion time, *RFA* is the number of requests for aid, and *NE* is the number of non-solved errors. Only the formula 8.2 is presented to the participants, the definition of the *TimeScore* variable is hidden. The name "TimeScore" suggests a bonus for fast task performances. Consequently, the participants are also motivated to rapidly perform the assembly task. Since the exact time scoring is hidden, they can not calculate the direct costs of non-solved errors and requests for aid. They just see the penalization weights and accordingly they will try to perform the assembly task without consulting the photographs.

8.3.2. Results and Analysis

The results of the first part of the exploitation questionnaire (see Figure 8.9) concerning the subjective assessment of the platform's help to learn the assembly task are illustrated in Figure 8.11. The results

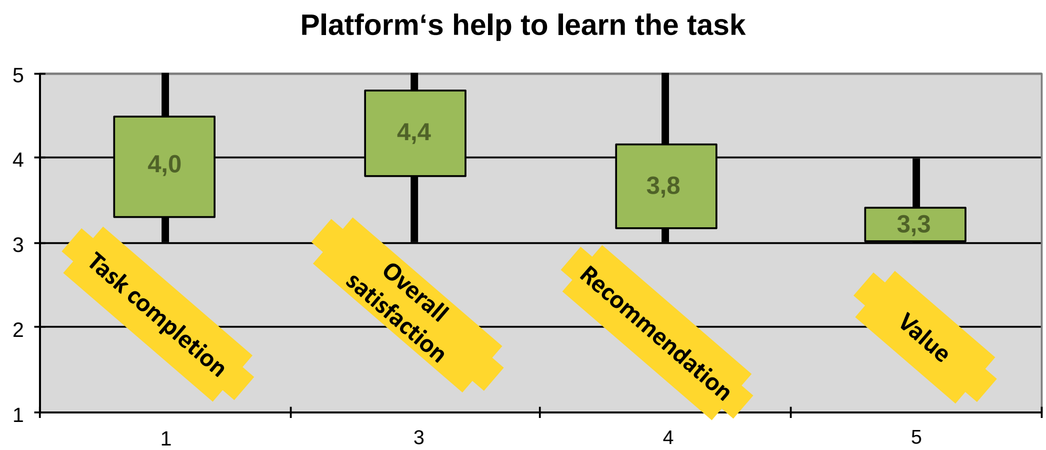


Figure 8.11.: Overview of the subjective assessment of the platform's help to learn the assembly task.

show, that the participants were overall very satisfied of how they were able to perform the assembly task in the evaluation session after training it using the AR-based training platform (average score of 4.33 of possible 5). The degree to which they completed the task was rated on average with 3.89, which is also a very good result as only one training cycle has been carried out. The average rating of the statement "I would recommend the platform for training the task I experienced" is 3.67, with a minimum of 3 and a maximum of 5. Hence, no one of the participants would definitely not recommend the platform for training the assembly task, while half of the participants would definitely recommend it. Furthermore, all ten participants stated, that training using the AR-based training platform quickly enhanced their skill level in the trained task (see Appendix C.2.4). Overall the platform has been assessed as a valuable training tool (average score is 3.22), even if the rating of the trainers for this question is higher. This discrepancy may be a result of the trainers' much higher experience and knowledge level in the field of technical training. As the trainers have a deeper understanding of the training needs, their assessment of the usefulness of the developed training platform differs from the technicians' rating. Beyond that, the trainers know the traditional training methods very well and accordingly the subjective comparison to these training methods inevitably affects the assessment of the developed training platform.

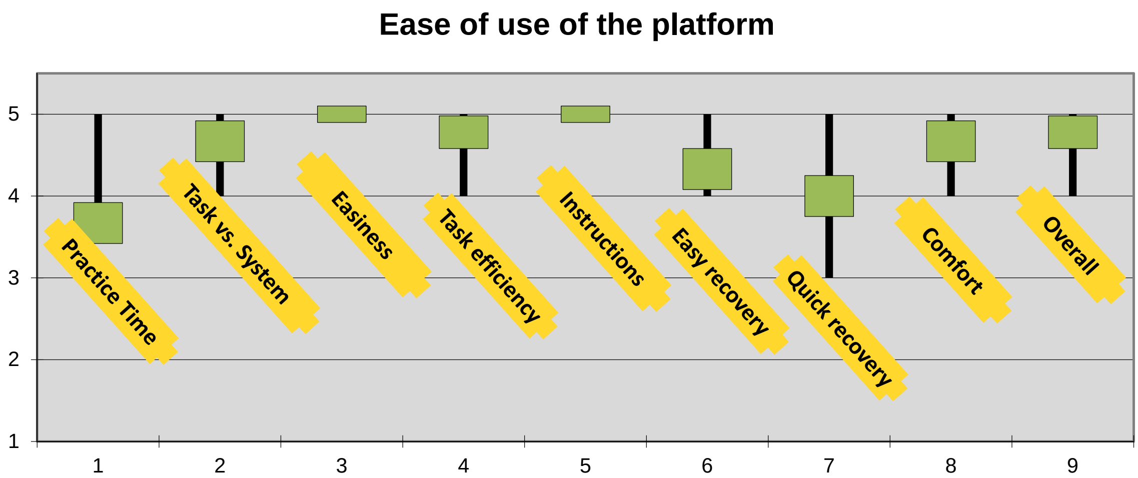


Figure 8.12.: Overview of the subjective assessment of the ease of use of the developed training platform.

An overview of the ratings concerning the ease of use of the training platform is presented in Figure 8.12. Nearly all mean scores are above 4.6 of possible 5.0. The easiness of accomplishing the assembly task using the AR-based training platform and the clarity of the provided instructions were even rated with highest possible score by all participants. The average scores for the error recovery possibilities (4.33 and 4.0) of the developed training platform are a slightly below the other ratings, but they are still good. We assume that the reason for this are the step switching possibilities of the training application, which only allows for switching to the direct precedent step and to the direct subsequent step, but not for skipping steps. So, for example, if an error made in an early step of the task is detected at a late stage of the performance of the task, the user has go stepwise back in the application until the failed step is reached (even if it is not required to physically undo all steps lying between). The only aspect that is assessed with an average score a bit lower than 4.0, namely 3.67, is the time required for practicing the assembly task using the developed training platform (i.e. the time required for performing one training cycle, in the following called training duration). However, Figure 8.13 shows that the average training duration of the AR group is not much higher than the one of the control group members. The training durations of the control group are a bit shorter, because some of the control group members decided not to physically perform the assembly task. The detailed ratings of all participants can be found in Appendix C.2.4.

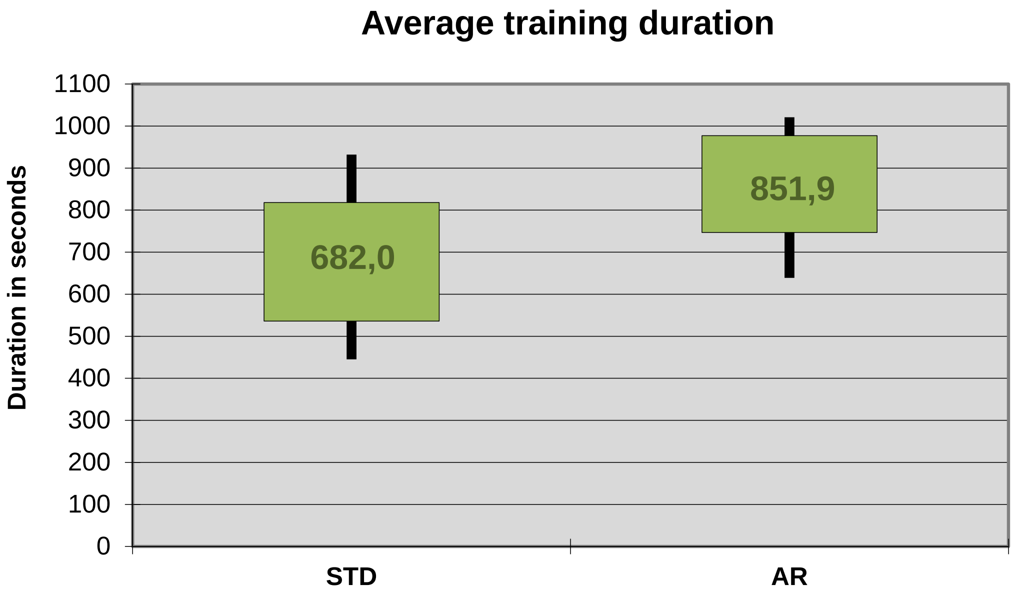


Figure 8.13.: Overview of the training durations of the AR group members (AR) and control group members (STD).

An overview of the measured performance times of the capability test is provided in Figure 8.14. The mean performance times of both groups are very close, also the distribution of the performance times are similar. During the capability test to non-solved errors were made, only three participants of the AR group failed one or two times, but could solve the errors on their own (see Appendix C.2.5). The initial skill level of all participants can be considered as advanced, since each of them could accomplish the *item* assembly task without any help. From this we can conclude, that the groups are homogeneous and that the allocation of the groups was chosen well. Figure 8.15 illustrate the average performance times of the participants in the evaluation test. The mean value of the control group is about 517 seconds (8min 37s), the AR group achieved an average value of 494 seconds (8min 14s). Hence, the mean performance time of the AR group is about 23 seconds (4.5%) below the average value of the control group, which is a good result if we consider that the selected training task is quite simple to perform and

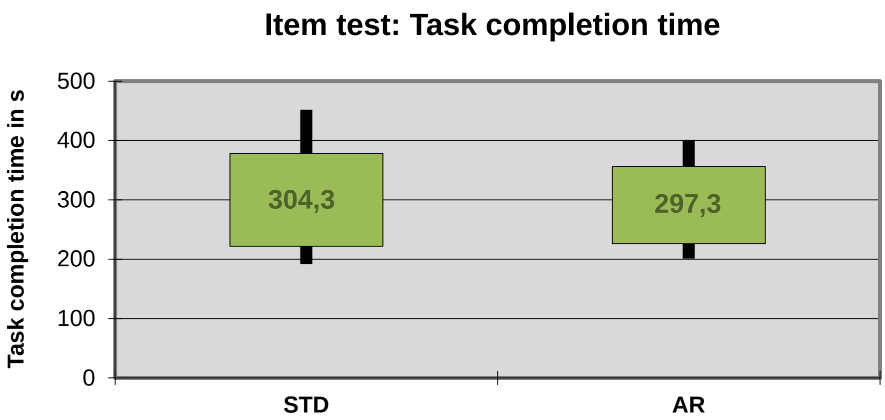


Figure 8.14.: Overview of the task completion times during the capability test (*item* test) of both groups.

understand for technicians. Mainly the control group members who decided not to physically perform the assembly task during training needed the longest time for accomplishing the evaluation test.

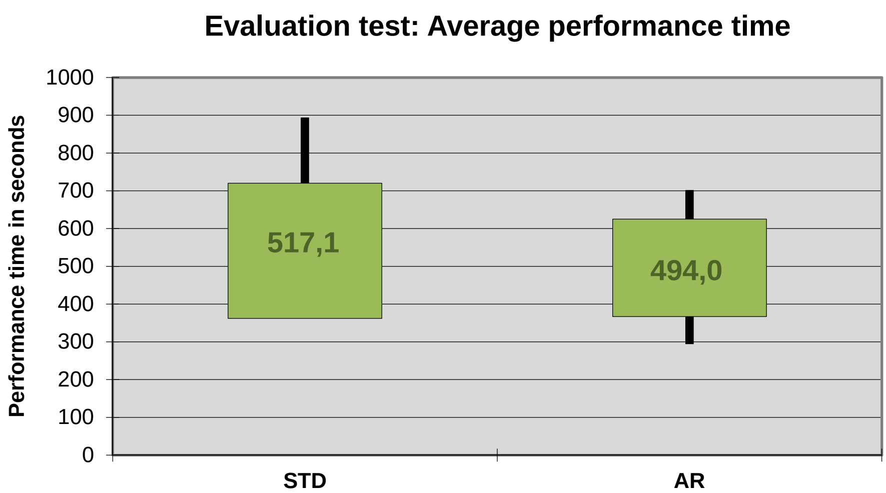


Figure 8.15.: Overview of the average performance times of the AR group (AR) and the control group (STD) in the evaluation test.

The error rates are summarized in Figure 8.16. The results show that the AR group not only performed faster, but also was able to perform the task with less errors. All in all, the control group performed 93.2% of actions correct and without any aids, while the AR group achieved a rate of 97.6%. The rate of errors that could be solved is the same for both groups, namely 1.2%. However, the rate of errors that could not be solved by the participants is much higher for the control group (5.6%, corresponds to 14 errors) than for the AR group (1.2%, corrsponds to 3 errors). Here, for both groups the errors—solved and non-solved—made during the performances consist mainly of wrong placements of components (see Figure 8.17). This shows that a clear indication of the location and the orientation to/in which components have to be attached is an important factor that should be taken into account when creating and selecting training material. As the AR group made less errors, the combination of an appropriate demonstration of the action (e.g. using text plus video), which is provided by the Adaptive Visual Aid content object, and a spatial information provided by the Adaptive Visual Aid pointer is useful to clarify how to place components.

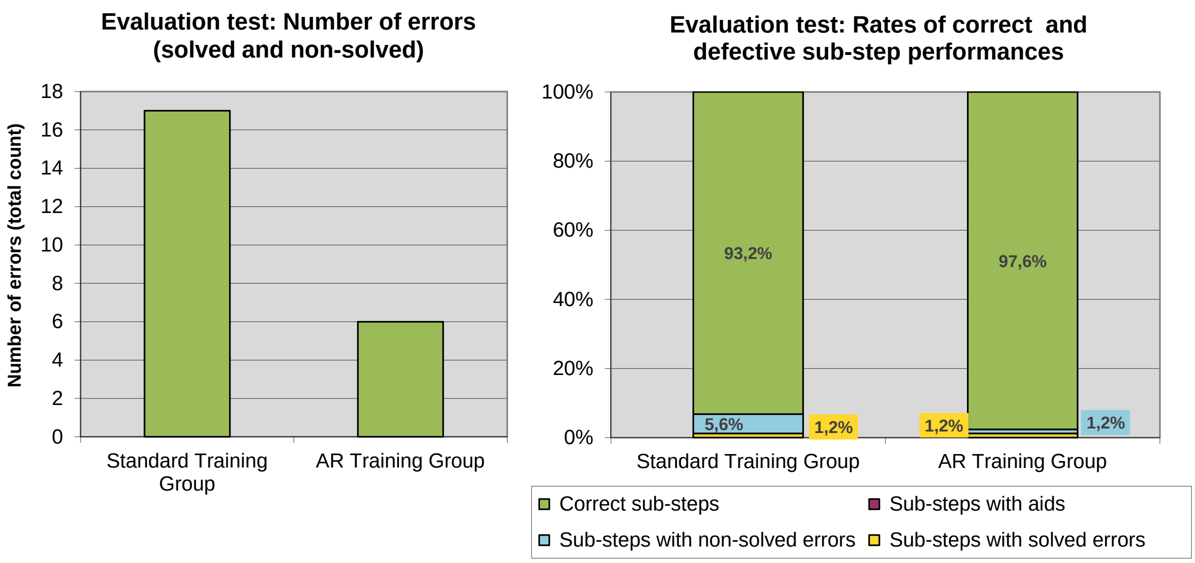


Figure 8.16.: The errors made by both groups in the evaluation test. Left: the total number of errors (solved and non-solved); right: the rates of correct and defective performances of the task.

An overview of the average performance scores of the AR group and the control group calculated using Formula 8.1 is given in Figure 8.18. The AR group not only performed faster and made less errors, but also reached a better average score (11.17) than the control group (8.58). This result is significant with $p<0.1$ ($p=0.07$). As illustrated in Figure 8.19, this applies not only for the average scores, but also for the individual scores of the participants: Without exception, each participant of the AR group outperformed his counterpart with the equivalent skill level in the control group in terms of performance scores.

Further details about the measurements taken in the skill transfer evaluation, including also the measurement results of the step completion times, can be found in the appendix (see Appendix C.2.5).

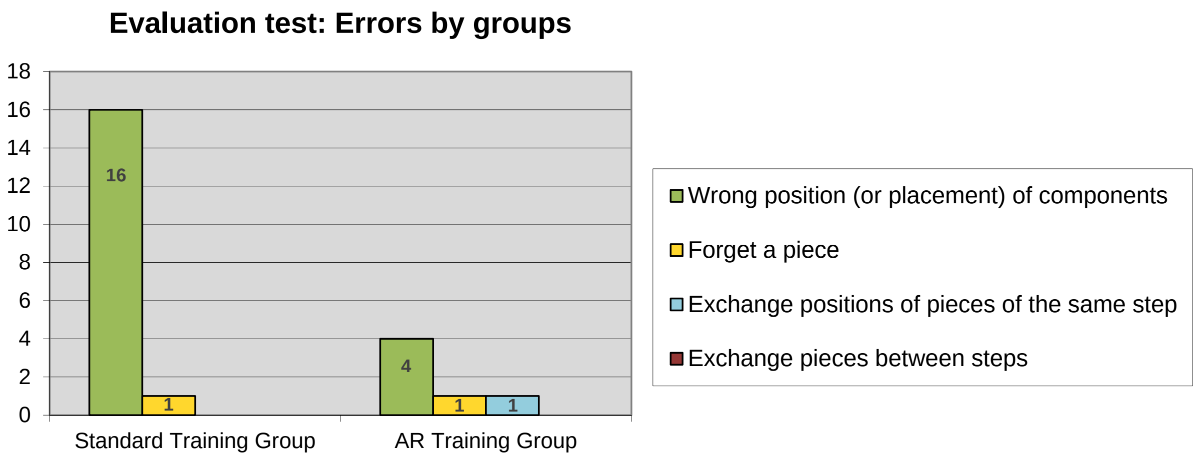


Figure 8.17.: Illustration of errors made by both groups in the evaluation test grouped by type of error.

8.3.3. Summary

To conclude, the results of the evaluation, which has been conducted in two stages, show that the developed training platform is well suited for the training of maintenance and assembly skills. This has been demonstrated based on two different approaches: On the one hand, there was the subjective assessments of trainers and trainees concerning the usability and functionality of the platform and the value of the applied training strategies. On the other hand, the performance of the trainees has been measured and analyzed. The usability of the platform has been assessed as very convenient and easy to understand by both trainers and trainees and has been praised in particular. The skills transfer evaluation further showed that the participants of the AR group were able to perform the evaluation test faster than the control group members and also with less errors. In addition, the average score—the most important factor—which has been calculated in order to assess the participant's skill level after the training is higher in the AR group than in the control group and is significant with $p<0.1$. The same applies for the individual scores of the participants. From this it can be concluded that the training of maintenance and assembly skills can be enhanced by using the implemented training platform and the proposed training strategies and accelerators.

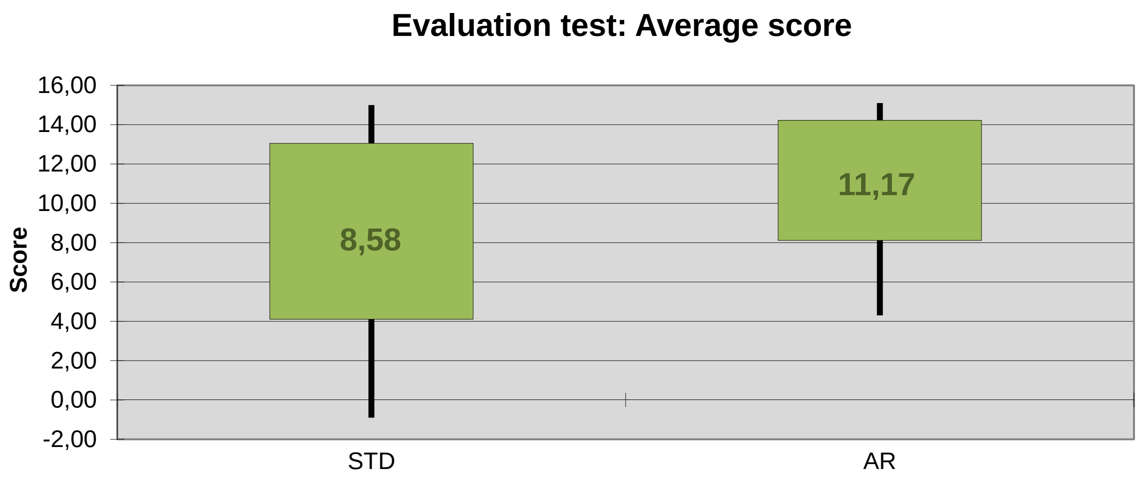


Figure 8.18.: Overview of the average performance scores of both groups in the evaluation test.

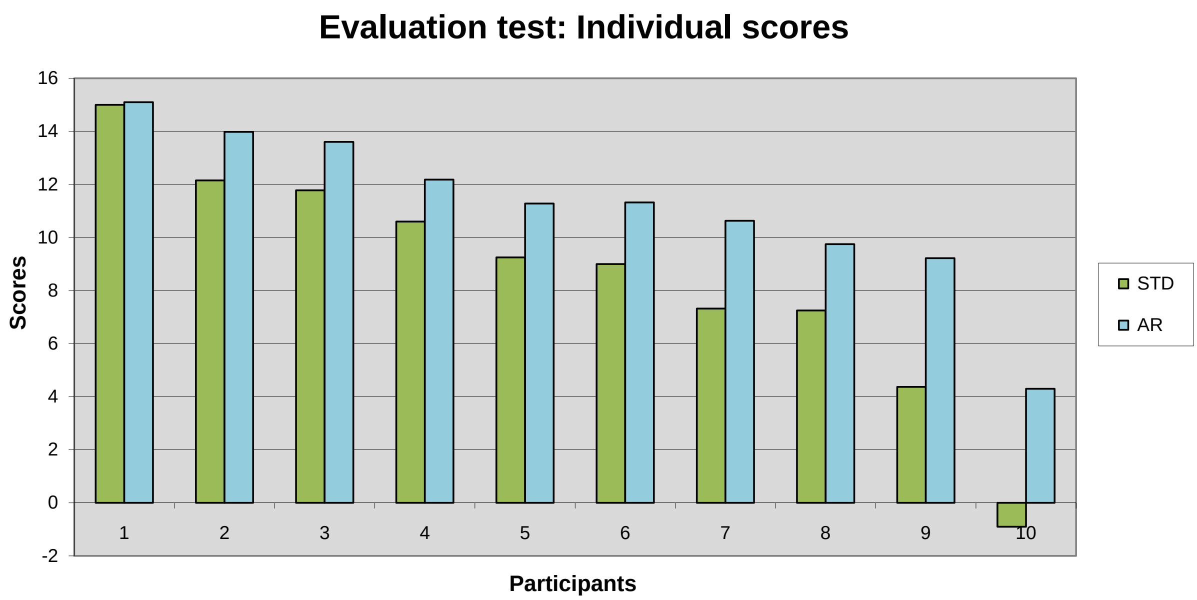


Figure 8.19.: Individual performance scores of the participants of both groups: each member of control group plotted against his counterpart of the AR group.

9. Conclusions

This chapter summarizes the content and results of the thesis. Furthermore, final conclusions will be drawn and an outlook for future work will be given.

9.1. Summary and Conclusion

In this thesis a novel concept and a platform for multimodal Augmented Reality-based training of maintenance and assembly skills have been developed, implemented, and evaluated. Following the introduction and analysis of existing AR-based training systems and skill training systems in general, relevant aspects of skill acquisition and training have been compiled, such as the decomposition of skills in the underlying sub-skills and approaches for the training and measuring of skills (Chapter 3). Subsequently, technologies that are relevant for AR-based skill training, such as capturing and rendering technologies, have been described and analyzed (Chapter 4).

Afterwards, we analyzed the use of Augmented Reality and multiple modalities for training in general. Here it was pointed out that the application of Augmented Reality technologies and the provision of multimodal feedback—and vibrotactile feedback in particular—have a great potential to enhance skill training. We also presented an approach for the development of training programs for multimodal training (Chapter 5.1). In this context we introduced the concept of accelerators, namely augmentation accelerators, variability accelerators, and simplification accelerators. These accelerators are variables that are introduced in the training process in order to facilitate and enhance training (Chapter 5.2). Since humans absorb most of the information through the eyes and hence vision dominates the modalities in the majority of cases, the visualization of information is vitally important for developing efficient training systems. Therefore, we presented how detailed information such as text can be displayed in interactive Mixed Reality environments. Furthermore, a novel concept for displaying location-dependent information in Augmented Reality environments was introduced which can compensate tracking imprecisions (Chapter 5.3). In this concept the pointer-content metaphor of annotating documents has been transferred to Augmented Reality environments. Therefore, so-called *Adaptive Visual Aids* (AVAs) have been defined. These AVAs consist of a tracking-dependent pointer object and a tracking-independent content object, both providing an adaptable level and type of information. Thus, the guidance level of Augmented Reality overlays in AR-based training applications can be easily controlled (Chapter 5.3.2). AVAs can be used to substitute the traditional Augmented Reality overlays (i.e. overlays in form of detailed 3D models or animations), which highly suffer from tracking inaccuracies.

In this thesis, also an design approach for the architecture of a multimodal AR-based training platform has been developed. This approach is not specific to the training of maintenance and assembly skills, but is a general design approach for multimodal training platforms. The proposed platform consists of three main building blocks:

- *Multimodal Capturing Controller:* This component handles the data delivered by different capturing systems (e.g. motion capture, camera tracking and object recognition, sound capturing, capturing of biometrics).
- *Interaction Processing and Application Module:* This module holds the application and workflow logic and processes the user input. The workflow block inside this module handles the whole workflow of the training task. Furthermore, it holds concepts for recognizing and interpreting patterns in the capturing data and for refining existing workflows based on captured data.
- *Multimodal Rendering Controller:* This block handles the rendering of the visual, haptic, and audio data.

Based on all these elaborated contents, a concept for the training of maintenance and assembly skills using a multimodal Augmented Reality-based platform has been developed, including the identification of necessary sub-skills, the training of the involved skills, and the design of a training program for the training of maintenance and assembly skills (Chapter 6). Since *procedural skills* are considered as the most important skills for maintenance and assembly operations, they have been discussed in detail, as well as appropriate methods for improving procedural skills. As a result, we defined training strategies and specific accelerators for the training of maintenance/assembly skills in general and procedural skills in particular. The proposed training strategies are the provision of *direct aids, indirect aids, and context aids* (Chapter 6.4.1). The specific accelerators—which implement the training strategies—are specified as follows:

- *Adaptive Visual Aids with Information on Request:* Instructions during training should be displayed using AVAs with a permanently visible AVA pointer and a AVA content providing additional information on user demand.
- *Device Display:* A Device Display providing information about the current sub-task (i.e. not only about the current step) and about the sub-goal to reach should be available in order to enhance the user's mental model building process. The information should contain the condition of the device before and after the sub-task.
- *Structure and Progress Information:* Information about the structure of the training task and the user's progress in the task should be provided. This also enhances the user's mental model building process.
- *Haptic (Vibrotactile) Hints:* Additional haptic or rather vibrotactile hints should be presented to the user during training. These hints can be used to clarify information that is difficult to visualize by using solely the available training material (e.g. presentation of rotating directions), to present subliminal information, or to provide error feedback (e.g. vibration hints when the user's grasps a wrong tool).

Also an exemplary training protocol for the training of maintenance skills has been developed. This includes the definition of task scenarios, the training procedure, and performance measures for assessing the user's performance of the training task (Chapter 6.4.2).

As for Augmented Reality there is no standard at all, the developed multimodal Augmented Reality training platform and the specified accelerators have been implemented based on the *X3D* ISO standard, which defines features that are very useful for Mixed Reality environments in general and Augmented Reality in particular. Implementing the training platform on the basis of a standard increases its sustainability and portability enormously. The implemented training platform has been evaluated at the food packaging manufacturer Sidel located in Parma (Italy) and compared to traditional training methods. Technical trainers and technicians working at the Sidel served as participants. The evaluation has been carried out in two stages: First, a *functionality and usability evaluation* and secondly, a *skill transfer evaluation* (Chapter 8). While in the first stage trainers served as participants, in the second evaluation stage technicians acted as subjects. The value of the training platform and the specified accelerators has been analyzed based on questionnaires about various aspects of the training platform, which the participants had to complete during the evaluation, and performance measurements taken during the evaluation. The results show that the developed training platform and the pursued training strategies are appropriate for the training of maintenance and assembly skills. Especially the usability of the platform has been assessed as very convenient. Also the pursued training strategies implemented through the specified accelerators were judged as very suitable for the training of maintenance and assembly skills. These subjective assessments have been affirmed by analyzing the performance measurements. The presentation of feedback using the haptic modality (via simple devices) in addition to the visual feedback has also been rated as valuable, in particular for presenting information that is difficult to visualize without ambiguity.

Concluding, Augmented Reality points out to be a good technology for training in the field of maintenance and assembly, as instructions or rather location-dependent information can be directly linked and/or attached to physical objects. Since machines to maintain usually contain a large number of similar components (e.g. screws, plugs, etc.), the provision of location-dependent information is vitally important for the training of maintenance operations. Another advantage is that AR-based training takes place in a real/physical environment with the physical devices involved in the training scenario. Thus, the trainees practice the physical performance of the task during training whereby the corresponding sensorimotor skills are trained. Furthermore, in Augmented Reality-based training sessions the availability of a trainer on-site is not mandatory.

The evaluation further showed that after one training session the skill level of technicians who trained with the developed training platform was higher than the skill level of those who performed the training session using traditional training methods. On the average, technicians who trained with the AR-based training platform made less errors and achieved better performance times, while the duration of one training cycle is only slightly higher than in traditional training. To conclude, it has been pointed out that the multimodal skill training platform and concept developed in this work enhance and improve training in the field of maintenance and assembly operations. Accordingly, the specified training strategies and accelerators—which implement the training strategies—can be considered as design guidelines for the development of Augmented Reality-based training systems for maintenance and assembly skills.

9.2. Future Work

Future work in the field of Augmented Reality-based training of maintenance and assembly operations should give special attention to the capturing and interpretation of the underlying skills. The recognition of the user's intention during the performance of single actions can be useful for providing well-directed and systematic feedback. Furthermore, if the maintenance skills of experts can be captured in large part, this data can be used to "train" the training system. That is, it can be used for example to refine the workflow and the instruction data. This way, the information and the instructions provided to the user during training can be improved.

Besides this, the integration of an additional remote component should be analyzed in future work. The presented concept of Adaptive Visual Aids allows for easily attaching annotations to objects (i.e. for attaching location-dependent information). Say that the trainee needs additional help when performing the training task and a trainer is not available on-site, but can remotely observe the trainee's performance of the task: in this case it may be useful for the trainee to receive additional location-directed hints from the remotely connected trainer in form of annotations that the trainer creates on-the-fly and attaches to the dedicated machine parts. For this reason it is advisable to evaluate if such a remote component for maintenance and assembly training improves training.

In addition, the exploration of different devices for providing haptic hints (i.e. simple, abstract haptic information) would be useful. It should also be explored, which kind of information and/or hints can be provided using such devices and which is the best way to present them. Being able to clearly present a large set of different types of information would enhance the possibilities of providing information via the haptic modality in Augmented Reality environments and, hence, would enable an intensified exploitation of the advantages of providing haptic stimuli during training.

A. Publications and Talks

The majority of the work described in this thesis has been peer-reviewed and presented at conferences. The thesis is partially based on the following publications and talks:

A.1. Publications

- Webel, Sabine; Bockholt, Ulrich; Engelke, Timo; Peveri, Matteo; Olbrich, Manuel; Preusche, Carsten: **Augmented Reality Training for Assembly and Maintenance Skills** To appear in: *Proceedings of the The International Conference of the SKILLS European Project*, 2011
- Gavish, Nirit; Gutierrez, Teresa; Webel, Sabine; Bockholt, Ulrich; Engelke, Timo; Peveri, Matteo; Olbrich, Manuel; Preusche, Carsten: **Design Guidelines for the Development of Virtual Reality and Augmented Reality Training Systems for Maintenance and Assembly Tasks** To appear in: *Proceedings of the The International Conference of the SKILLS European Project*, 2011
- Webel, Sabine: **Augmented Reality in Skill Training.** In: *Skill Learning and Virtual Environments: Skills Summer School 2011 (Information Society Technologies - IST)*, 2011
- Webel, Sabine; Bockholt, Ulrich: **Design Criteria for AR-Training Systems for Maintenance and Assembly Tasks.** In: *Virtual and Mixed Reality - New Trends: Part I: International Conference, Virtual and Mixed Reality 2011 (LNCS 6773)*, 2011
- Webel, Sabine; Bockholt, Ulrich; Engelke, Timo; Gavish, Nirit; Tecchia, Franco: **Design Recommendations for Augmented Reality based Training of Maintenance Skills.** In: *Recent Trends of Mobile Collaborative Augmented Reality Systems*, Springer-Verlag, 2011 - ISBN 978-1-4419-9844-6
- Engelke, Timo; Webel, Sabine; Gavish, Nirit: **Generating Vision based Lego Augmented Reality Training and Evaluation Systems.** In: *Proceedings of the Ninth IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2010
- Konietschke, Rainer; Tobergte, Andreas; Preusche, Carsten.; Tripicchio, Paolo; Ruffaldi, Emanuele; Webel, Sabine; Bockholt, Ulrich: **A Multimodal Training Platform for Minimally Invasive Robotic Surgery** In: *Proceedings of the 19th IEEE International Symposium in Robot and Human Interactive Communication (Ro-Man): Multimodal Interfaces for Capturing and Transfer of Skills*, 2010

- Engelke, Timo; Webel, Sabine; Bockholt, Ulrich; Wuest, Harald: **Towards Automatic Generation of AR-Training Applications and Workflow Descriptions** In: *Proceedings of the 19th IEEE International Symposium in Robot and Human Interactive Communication (Ro-Man): Multimodal Interfaces for Capturing and Transfer of Skills*, 2010
- Webel, Sabine; Engelke, Timo; Bockholt, Ulrich; Tecchia, Franco; Preusche, Carsten; Gavish, Nirit: **An AR Training Platform for Skills Transfer involved in Industrial Maintenance and Assembly Tasks** *Research Demonstration at IEEE Virtual Reality (IEEE VR)*, 2010
- Jung, Yvonne; Webel, Sabine; Olbrich, Manuel; Drevensek, Timm; Franke, Tobias; Roth, Marcus; Fellner, Dieter W.: **Interactive Textures as Spatial User Interfaces in X3D** In: *ACM SIGGRAPH: Proceedings of the 15th International Conference on 3D Web Technology (Web3D)*, 2010
- Webel, Sabine; Staykova, Yana; Bockholt, Ulrich: **Towards Workflow Acquisition of Assembly Skills using Hidden Markov Models** In: *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2009
- Preusche, Carsten; Konietschke, Rainer; Tobergte, Andreas; Ruffaldi, Emanuele; Webel, Sabine; Bockholt, Ulrich: **Development of a Multimodal Skills Trainer for Minimally Invasive Telerobotic Surgery** In: *Proceedings of the International Conference on Multimodal Interfaces for Skills Transfer (SKILLS - Enaction on SKILLS)*, 2009
- Gavish, Nirit; Engelke, Timo; Webel, Sabine; Bockholt, Ulrich: **Enhancing Trainee's Mental Model During AR Assembly Training** In: *Proceedings of the International Conference on Multimodal Interfaces for Skills Transfer (SKILLS - Enaction on SKILLS)*, 2009
- Casado, Sara; Engelke, Timo; Gavish, Nirit; Gutiérrez, Teresa; Rodríguez, Jorge; Sánchez, Emilio José; Tecchia, Franco; Webel, Sabine: **A Training System for Skills Transfer Involved in Industrial Maintenance and Assembly Tasks** In: *Proceedings of the International Conference on Multimodal Interfaces for Skills Transfer (SKILLS - Enaction on SKILLS)*, 2009
- Bockholt, Ulrich; Webel, Sabine; Engelke, Timo; Graf, Holger; Behr, Johannes: **Virtual and Augmented Reality Training Platforms for Transfer of Skills** In: *Proceedings of the International Conference on Multimodal Interfaces for Skills Transfer (SKILLS - Enaction on SKILLS)*, 2009
- Jung, Yvonne; Keil, Jens; Behr, Johannes; Webel, Sabine; Zöllner, Michael; Engelke, Timo; Wuest, Harald; Becker, Mario: **Adapting X3D for Multi-touch Environments** In: *ACM SIGGRAPH: Proceedings of the 13th International Symposium on 3D Web Technology (Web3D)*, 2008
- Webel, Sabine: **Annotationen in immersiven Umgebungen : Interaktionskonzepte und Paradigmen** VDM Verlag Dr. Müller, 2008 - ISBN 978-3-639-05627-3
- Webel, Sabine; Keil, Jens; Zöllner, Michael: **Multi-touch Gestural Interaction in X3D Using Hidden Markov Models** In: *ACM SIGCHI: Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST)*, 2008

- Webel, Sabine; Becker, Mario; Stricker, Didier; Wuest, Harald: **Identifying Differences Between CAD and Physical Mock-ups Using AR** In: *Proceedings of the Sixth IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2007
- Behr, Johannes; Dähne, Patrick; Jung, Yvonne; Webel, Sabine: **Beyond the Web Browser - X3D and Immersive VR** In: *IEEE Virtual Reality (IEEE VR): Proceedings of the IEEE Symposium on 3D User Interfaces*, 2007
- Jung, Yvonne; Weber, Christine; Webel, Sabine; Knöpfle, Christian: **Exploration einer virtuellen Stratigraphie mit Hilfe von Verfahren des Interaktiven Designreviews** In: *4. GI-Workshop der Fachgruppe Virtuelle Realität und Augmented Reality*, 2007

A.2. Talks

- Webel, Sabine: **Augmented Reality in Skill Training** At: *SKILLS Summer School 2011: Skill learning and Virtual Environments*, Arezzo (Italy), July 25-30, 2011
- Webel, Sabine: **AR-basiertes multimodales Training von Wartungs- und Montagefähigkeiten** At: *DLR - German Aerospace Center*, Munich (Germany), 2011
- Webel, Sabine; Bockholt, Ulrich: **Immersive Annotation - Ein Annotationssystem für immersive Umgebungen** At: *10. Fraunhofer IFF-Wissenschaftstage*, Magdeburg (Germany), 2007

B. Curriculum Vitae

Personal Data

Name	Sabine Webel
Birth date & place	03.02.1981 in Worms, Germany
Nationality	german

Education

11/2005	Graduation in Computer Science at Technische Universität Darmstadt, in Darmstadt, Germany
10/2000 – 11/2005	Study at Technische Universität Darmstadt in Darmstadt, Germany
09/2002 – 02/2003	Exchange Student at Politecnico di Torino in Turin, Italy

Work Experience

01/2006 – today	Research associate, Dep. Virtual and Augmented Reality, Fraunhofer Institute for Computer Graphics Research (IGD), Darmstadt, Germany
11/2008 – 03/2009	DAAD Scholarship: Guest Researcher at Universidade Federal do Espírito Santo (UFES), Vitória, Brazil

C. Evaluation Documents

C.1. Functionality and Usability Evaluation Documents

“Usability Evaluation“ Questionnaire

		SCALE		T1	T2	T3	T4	AVG
1	I was completely concentrated more on the task activities rather than on how to use the system used during the training session.	1	7	6	6	6	6	6
2	I felt totally comfortable during the training session.	1	7	6	5	7	6	6
3	The interaction with the IMA-AR platform was very easy.	1	7	6	7	6	6	6,25
4	I could ask for information (step information, mental model display) very easy within the application	1	7	6	6	6	7	6,25
5	I did not have any problem to understand how to use the platform	1	7	7	7	7	7	7
6	I think that I am very proficient user using the IMA-AR platform after the training session	1	7	6	6	7	7	6,5
7	From the usability point of view, Could you give a grade to the platform in overall?	1	10	8	4	9	9	7,5
8	From the interaction point of view, which elements of the system did you not like? which elements of the system were more difficult to understand?, Do you suggest any modification or improvement?	N/A		- Sometimes lose tracking when covering accidentally the rectangle				

“Functionality Evaluation“ Questionnaire

		SCALE		T1	T2	T3	T4	AVG
1	The information provided by the platform via displayed messages (step description, instructions,,,) was enough to understand the task.	1	7	4	5	6	6	5,25
2	What information would you modify/remove/add it?	N/A		- Overview of the filling process				
3	The visualisation of the different operations was enough for leaning the task?	1	7	5	5	6	6	5,5
4	Which steps would you modify/remove/add?	N/A						
5	Is there any critical information of the task missing?	N/A					no	
6	Could you rate the general visualisation utilities: spatial information (i.e. pulsing circles), step information, captions, ...?	1	10	6	7	8	8	7,25
7	Could you rate the overview strategy?	1	10	10	7	8	8	8,25
8	Could you rate the direct aids strategy? (spatial information, i.e. pulsing circles)	1	10	6	4	8	7	6,25
9	Could you rate the indirect aids strategy? (step information, mental model display)	1	10	7	8	10	7	8
10	Could you rate the context aids strategy? (progress bars)	1	10	6	7	6	8	6,75
11	Could you rate the haptic hints strategy?	1	10	6	3	2	8	4,75
12	Could you rate the playback/trainer-trainee based strategies?	1	10	6	8	10	8	8
13	Could you rate additional features such as: haptic hints, entire model visualisation, configuration possibilities of the demonstrator (i.e. allow mental model, enable haptic hints, use captions always/fade out) ...?	1	10	4	6	7	8	6,25
14	From the functionality point of view, could you give a grade to the platform in overall?	1	10	6	6	9	8	7,25
15	What percentage of the task do you consider that you have learnt?	%		90%	70%	80%	10%	62,50%
16	What grade would you give to the IMA-AR platform as learning system?	1	10	8	7	8	5	7
17	What features of the system did you like more?	N/A		- Sudden availability of the system - Interaction				
18	What features of the system did not you like?, Do you suggest any modification or improvement?	N/A		- when the tracking is lost a popup message that warn about that would be helpful - the system is good only for simple operations				
19	Do you have any additional comment?	N/A						
20	Do you see any advantage of this training system with respect to the traditional one?	N/A		- the chance to carry out the activity when I have time - not affected by errors like humans - no trainer needed				
21	Do you see any disadvantage of this training system with respect to the traditional one?	N/A		- No info regarding the whole filling process - No indication In case of minor errors				

C.2. Skill Transfer Evaluation Documents

C.2.1. Evaluation Protocol

Skill Transfer Evaluation

–

Evaluation Protocol

Augmented Reality-based Training System for Maintenance and Assembly Skills

Teresa Gutierrez[*], Sabine Webel[+], Matteo Peveri[**], Nirit Gavish[++]

[*] teresa.gutierrez@tecnalia.com (TECNALIA)
[+] sabine.webel@igd.fraunhofer.de (Fraunhofer IGD)
[**] Matteo.Peveri@sidel.com (Sidel S.p.A.)
[++] niritg@techunix.technion.ac.il (TECHNION)

Table of Contents

1 General

The transfer evaluation of the selected assembly task will be performed at Sidel with Sidel's technicians. The aim is to evaluate the use of the Augmented Reality-based training platform for the purpose of training in comparison with traditional training. Trainees will have training using either original training videos in combination with the real actuator (traditional training methods) or with the developed AR-based training platform. The training will be followed by a break of several hours and a test on the real actuator.

The evaluation will involve two main groups:

1. **<u>GROUP 1 – Control-AR</u>:**
   Training: 1 session of assembling the real actuator with the use of the videos.
   Test: Assembling of the real actuator with the use of photographs only is necessary.

2. **<u>GROUP 2 – AR</u>:**
   Training: 1 session of assembling the real actuator with the AR platform.
   Test: Assembling of the real actuator with the use of photographs only is necessary.

During the test, the evaluator will fill in a score-sheet with the relevant measures (times, number of errors, number of aids, etc.) and user's comments/problems that can be useful to analyse and interpret the results. In addition, a log file of the AR platform provides information about the trainee's interaction with the training platform during the task training. The trainee's actions will be recorded using a video camera for further analysis.

2 Design

The experiment will follow a between-participants design (each trainee will be assigned to only one group). For the assignment of the users to the groups, first they will have to fill in a demographic questionnaire whose answers will be used to distribute the users along the four groups in a homogeneous way.

3 Participants

About 20 technicians from Sidel will serve as participants. Participants will have at least 2 years of experience on field assembly/disassembly operations. Trainees will not receive any payment for their participation, but the best performers will be published. All of them will be competent user of personal computers and hi-tech in general.

The previous knowledge of participants will be tested through a simple assembly task performed with a specific Item set (capability test), just before starting the test.

4 Material and Set-up

The evaluation will be performed in Sidel facilities, in the Demonstrators room at Training Center, floor 2.

The evaluators will need the following material to conduct the evaluation:

- Item set for capability test
- Original training/instruction videos for the Control-AR group
- Books of photographs to be used during the real task
- A video camera to record trainees' actions with the actuator.
- Copy of the protocol
- Handouts for the platform familiarization session
- Copy of questionnaires for the trainees (demographic questionnaire and subjective evaluation questionnaire)
- Copy of a pre-prepared score-sheet where the evaluators will write down comments and record measures during the evaluation and the use of the photographs

Material for the AR platform

- Actuator (motorized valve), electro-mechanic type. The necessary parts should be removed before each assembly and should be put in front of the trainee in a pre-defined order.
- Tools which are put in front of the trainee in a predefined order.

<u>Important:</u>
The actuator and the tools will not be visible for the trainees until the need to use them!

5 Procedure

Text in italics is information for the evaluator and should NOT be read to the participant.

	Text presented like this, with the speech balloon icon, should be read out to the participant.

Each group will go through 6 different steps: introducing the task and the evaluation; introducing the training platform; familiarization with the training platform; training; test; questionnaires and technical capability test.

5.1 Capability Test

Before the training scenario the user will be requested to complete a simple assembly task using a set of Items.

The time will be recorded in order to complete the whole assembly. Starting from the situation in the picture below:

The trainee has to obtain:

The sequence of steps should be this:

 1	 2
 3	 4
 5	

5.2 Introduction of Task and Evaluation

	We are working in a European project, the SKILLS project, whose main goal is to develop new multimodal environments for improving the human skills transfer. Specifically, we are working in developing a new system for the transfer of skills involved in assembly and maintenance tasks. We have implemented different learning strategies, and we would like to evaluate some of them. You are going to participate in an experiment in which you will be required to learn how assemble some parts of an actuator. After the training you will be required to perform the real assembly as fast as possible and without errors. The names of the best performers will be published. The training part will be conducted now, and the test will be on the afternoon. The task consists in assembly a group of the electro-mechanic valve, the actuator. It starts with the assembly of the drive chain. Then you need to proceed with placing the electronic board, connecting the sensors and the drive, placing the cover, and finally mounting the clamp.

5.3 Introduction of the Training Platform

For the Control-AR group

	First you will be given a video showing the assembly task you will have to perform and then you will perform the task once with the help of a power point presentation on which task explanations is available.

For the AR group

	You will use a tablet PC with touch-screen. Put on the haptic bracelet on your right forearm/wrist and turn it on. You should sense six short vibration stimuli at different locations around your wrist.  Launch the start_IMA.bat and the configuration screen will appear. Your

trainer will configure and initiate the application.

Initialize the tracking by placing the marker (see Figure 1 below) in the camera image. Once it is initialized (the virtual valve appears on the in superimposition with the real valve) you should move the camera a little bit around in order to stabilize the tracking. Now the marker does not necessarily be in the camera image. Every time the tracking gets lost, you should move the camera to place the marker in the image in order to re-initialize the tracking.

Figure 1: The marker for initializing the tracking of the AR training platform.

5.4 Familiarization with the training platform

5.4.1 Control-AR group: Familiarization with the traditional training methods

No familiarization is necessary. Trainees are already familiar with this training scenario.

5.4.2 AR group: Familiarization with the AR-based training system

When the training application is initiated, you can go step-by-step through the training task. You can switch to the next/previous step by touch the arrow right/arrow left button (see Figure 2 and 3).

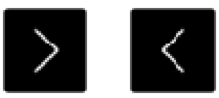


Figure 2: next step-button (left) and previous step-button (right)

	At the top of the window there is a progress-bar showing your progress in the training task (see Figure 3).

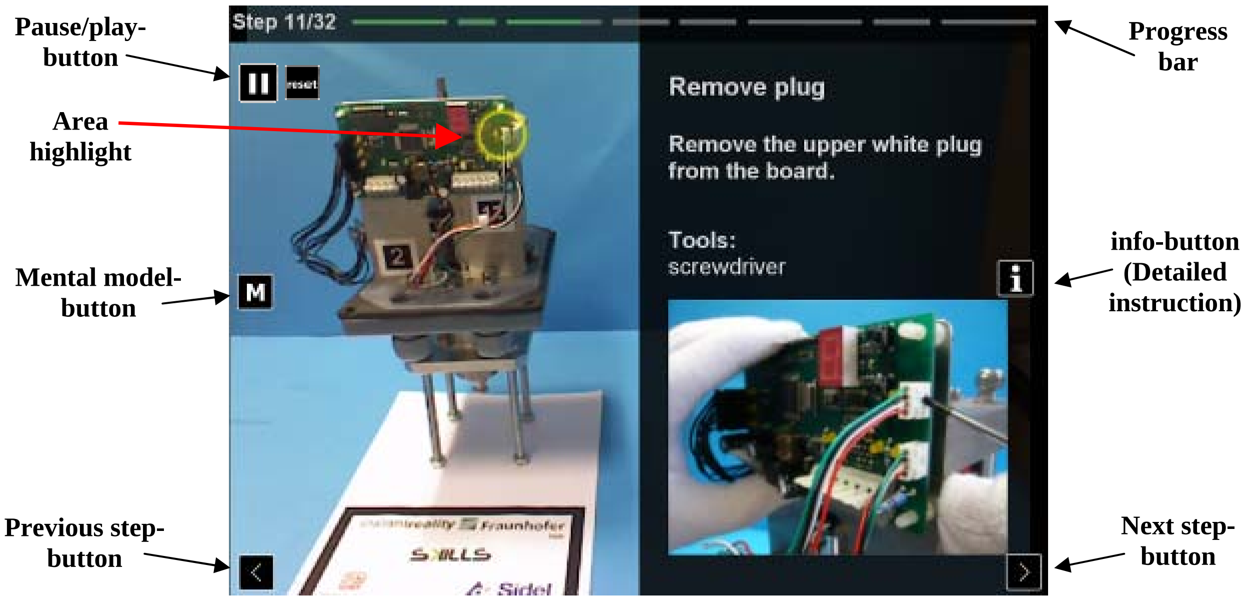
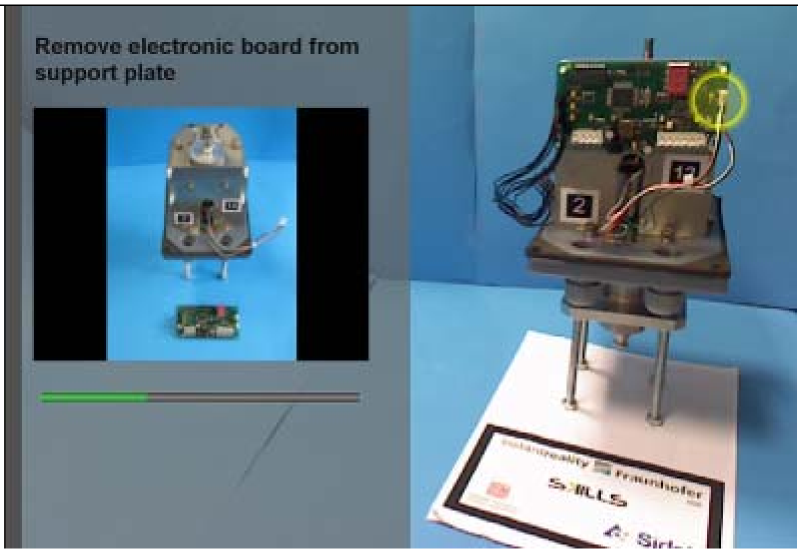


Figure 3: Information, instruction and button elements

	Beginning from the first step information about the current step is provided. A yellow pulsing circle/ellipse overlaid onto the camera image highlights the area of interest for the performance of the next step (see Figure 3). By touching the info-button (see Figure 3) the detailed instruction is displayed (see Figure 3). The instruction contains also information about the tools to use. The instruction can be hidden by touching the plane containing the instruction. If available, you get also vibrotactile hints (i.e. vibration stimuli at your arm indicating a movement to perform).

	By touching the *mental model-button*, information about the "mental model" is provided. This means, that it is shown how to reach the next sub-goal of the task. This mental model information can be deactivated by touching the plane containing the information.

	 The progress-bar inside this "mental model" information shows your progress inside the group steps belonging together to reach the sub-goal.

<u>Instructions for the evaluator:</u>

Now start the application and configure the start screen, by setting

- ***Mode:*** *"assembly"*
- ***Visual Aids / Augmentation:*** *next step and mental model*
- ***Additional Aids:*** *enable vibrotactile feedback*

	In step 0 a 3D model of the valve is overlaid onto the camera image of the real valve. This shows you, how the device will appear after a successful assembly task. Switch to the first step by touching the next step-button (see Figure 3).

	The yellow pulsing circle shows you the area of interest for performing the next step. Now you can ask for detailed step information (that is, a detailed instruction) and for the "mental model" (that is, the next sub-goal to achieve). Please try now to access the "step information" and the "mental model" information.

If the trainee is not able to initialize the tracking and work with it, show him how to move the camera in order to initialize the tracking using the marker and to get a stable tracking. Now ask the trainee to activate and deactivate

- *the "mental model display"*
- *the "step information"*

If he does not know how to do it, show him the buttons he has to touch.

Now tell the trainee to freeze the current view (i.e. the current camera image) and afterwards to continue the live camera view.
If he does not know how to do it, show him the pause/play button he has to touch.

When you have understood what to do, perform the step at the real valve. If you need information you can always ask the system by touching the corresponding buttons. You can also always freeze the current view using the pause/play-button. The image on the button changes when you touch it. When you are freezing the view, the image changes to a play-icon (arrow right, see Figure 4 right). This indicates that by touching the button, the camera image continues "playing". When the live view continues, the image changes again to a pause-icon (see Figure 4 left), which indicates that by touching it you can pause the camera view again.

Figure 4: pause/play-button: pause-icon (left) and play-icon (right)

After performing the step you can switch to the next step. The handling of the system is identical for each step.

More familiarization with the system is not required.

5.5 Training in the task

The training should be recorded with a camera.

Once you have learnt to use the training platform, now we will start the test. Using this platform you must learn a real assembly task: assemble some parts of an actuator. This task consists of 25 steps grouped in 6 sub-tasks

1) Roller → Place the Belt Roller
2) Sensors → Place the Two Level Sensors
3) Support Plate → Fix the Support Plate into the structure
4) Electronic Board → Place the Electronic Board
5) Actuator Cover → Place the Actuator Cover
6) Clamp; Place the Clamp into the actuator stem

5.5.1 Control-AR group: Training in the task using traditional training methods

You will now perform the task with the help of the videos and you will have the occasion to try the steps once on the real parts.
Remember that your goal is to learn the task and you will be tested on performing it by yourself later on.

5.5.2 AR group: Training in the task with the AR-based training platform

	You will now perform the task with the help of the AR. Remember that your goal is to learn the task and you will be tested on performing it by yourself later on.

5.5.3 Final instructions for all groups

For all groups – after the training

	Thank you for your participation so far. Please come back at XXX to undertake the real task. But before leaving, could you answer me a few questions about your experience with the training system.

Ask the user the questions of the questionnaire "subjective evaluation"

5.6 Test

	Please perform the assembly task in which you have trained before. At the end you should have the actuator assembled without any error and being functional (i.e. all the pieces must be well placed and fixed). You should perform the task by yourself without any help. But if for any reason you can not continue with the task, you can ask me help. For each one of the steps of the task I have a picture with information about the step. If you want me to show you the picture of any specific step, please tell me the step and I will show you its corresponding picture. But if you do not know the step, I will show you the picture for the next step. With these pictures, you must find the information you need in order to be able to continue with the task. But, please ask me "help" just only if it strictly necessary to continue the task, since the use of these pictures will be a penalization in the final result. Your score will be based on the number of non-solved errors at the end of the task and the number of pictures used during the task. Score = TimeScore - 1* no. of unsolved errors - 2* (no. of pictures used) The parameter TimeScore will be defined on the basis of the average time completion of the task. Better score means better performances. The names of the best performers will be published.

C.2.2. Familiarization Document

Skill Transfer Evaluation

–

Familiarization Document

Augmented Reality-based Training System for Maintenance and Assembly Skills

Teresa Gutierrez[*], Sabine Webel[+], Matteo Peveri[**], Nirit Gavish[++]

[*] teresa.gutierrez@tecnalia.com (TECNALIA)
[+] sabine.webel@igd.fraunhofer.de (Fraunhofer IGD)
[**] Matteo.Peveri@sidel.com (Sidel S.p.A.)
[++] niritg@techunix.technion.ac.il (TECHNION)

Table of Contents

1 Introduction of the Training Platform

For the Control-AR group

First you will be given a video showing the assembly task you will have to perform and then you will perform the task once with the help of a power point presentation on which task explanations is available.

For the AR group

You will use a tablet PC with touch-screen. Put on the haptic bracelet on your right forearm/wrist and turn it on. You should sense six short vibration stimuli at different locations around your wrist.

Launch the start_IMA.bat and the configuration screen will appear. Your trainer will configure and initiate the application.

Initialize the tracking by placing the marker (see Figure 1 below) in the camera image. Once it is initialized (the virtual valve appears on the in superimposition with the real valve) you should move the camera a little bit around in order to stabilize the tracking. Now the marker does not necessarily be in the camera image. Every time the tracking gets lost, you should move the camera to place the marker in the image in order to re-initialize the tracking.

Figure 1: The marker for initializing the tracking of the AR training platform.

2 Familiarization with the training platform

2.1 Familiarization with the traditional training for the Control-AR group

No familiarization is necessary. Trainees are already familiar with this training scenario.

2.2 Familiarization with the AR-based training platform for the AR group

When the training application is initiated, you can go step-by-step through the training task. You can switch to the next/previous step by touch the arrow right/arrow left button (see Figure 2 and 3).

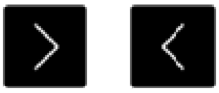


Figure 2: next step-button (left) and previous step-button (right)

At the top of the window there is a progress-bar showing your progress in the training task (see Figure 3).

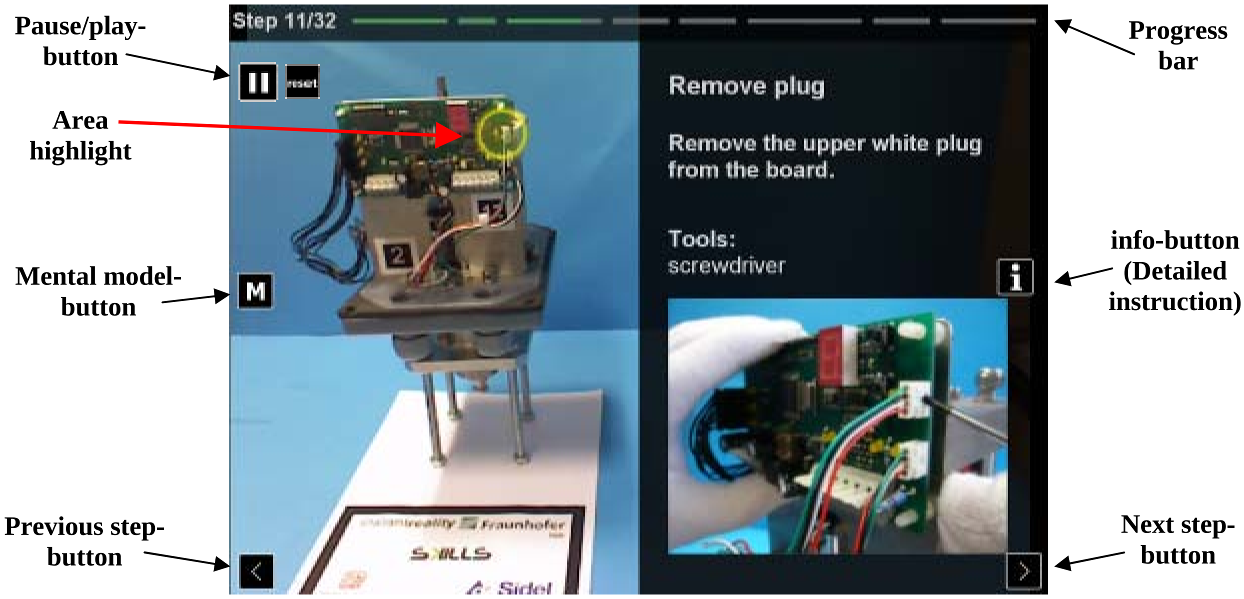
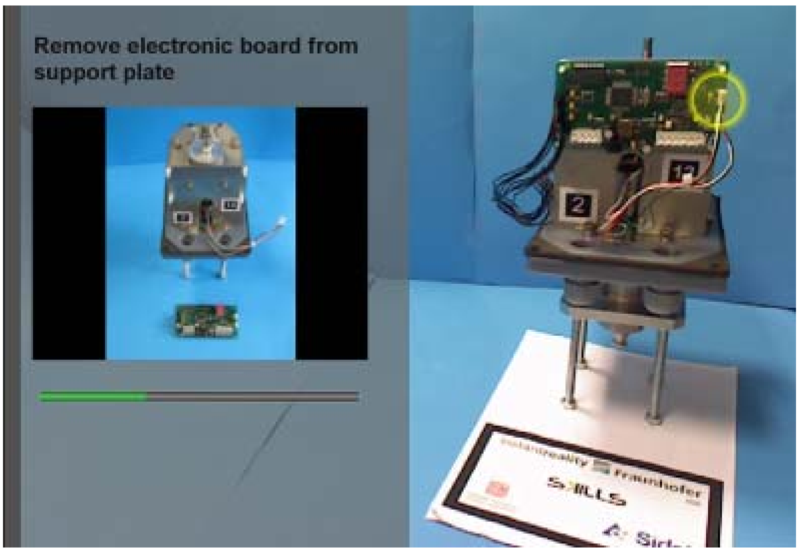


Figure 3: Information, instruction and button elements

Beginning from the first step information about the current step is provided. A yellow pulsing circle/ellipse overlaid onto the camera image highlights the area of interest for the performance of the next step (see Figure 3). By touching the info-button (see Figure 3) the detailed instruction is displayed (see Figure 3). The instruction contains also information about the tools to use. The instruction can be hidden by touching the plane containing the instruction. If available, you get also vibrotactile hints (i.e. vibration stimuli at your arm indicating a movement to perform).

By touching the *mental model-button,* information about the "mental model" is provided. This means, that it is shown how to reach the next sub-goal of the task. This mental model information can be deactivated by touching the plane containing the information.



	The progress-bar inside this "mental model" information shows your progress inside the group steps belonging together to reach the sub-goal.

<u>Instructions for the evaluator:</u>

Now start the application and configure the start screen, by setting

- ***Mode:** "assembly"*
- ***Visual Aids / Augmentation:** next step and mental model*
- ***Additional Aids:** enable vibrotactile feedback*

	In step 0 a 3D model of the valve is overlaid onto the camera image of the real valve. This shows you, how the device will appear after a successful assembly task. Switch to the first step by touching the next step-button (see Figure 3).

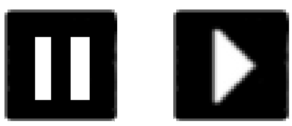
	The yellow pulsing circle shows you the area of interest for performing the next step. Now you can ask for detailed step information (that is, a detailed instruction) and for the "mental model" (that is, the next sub-goal to achieve). Please try now to access the "step information" and the "mental model" information.

If the trainee is not able to initialize the tracking and work with it, show him how to move the camera in order to initialize the tracking using the marker and to get a stable tracking. Now ask the trainee to activate and deactivate

- *the "mental model display"*
- *the "step information"*

If he does not know how to do it, show him the buttons he has to touch.

Now tell the trainee to freeze the current view (i.e. the current camera image) and afterwards to continue the live camera view.
If he does not know how to do it, show him the pause/play button he has to touch.

	When you have understood what to do, perform the step at the real valve. If you need information you can always ask the system by touching the corresponding buttons. You can also always freeze the current view using the pause/play-button. The image on the button changes when you touch it. When you are freezing the view, the image changes to a play-icon (arrow right, see Figure 4 right). This indicates that by touching the button, the camera image continues "playing". When the live view continues, the image changes again to a pause-icon (see Figure 4 left), which indicates that by touching it you can pause the camera view again.  **Figure 4: pause/play-button:** pause-icon (left) and play-icon (right)

After performing the step you can switch to the next step. The handling of the system is identical for each step.

More familiarization with the system is not required.

3 Training in the task

The training should be recorded with a camera.

Once you have learnt to use the training platform, now we will start the test. Using this platform you must learn a real assembly task: assemble some parts of an actuator. This task consists of 25 steps grouped in 6 sub-tasks

1) Roller → Place the Belt Roller
2) Sensors → Place the Two Level Sensors
3) Support Plate → Fix the Support Plate into the structure
4) Electronic Board → Place the Electronic Board
5) Actuator Cover → Place the Actuator Cover
6) Clamp; Place the Clamp into the actuator stem

3.1 Control-AR group: Training in the task using traditional training methods

You will now perform the task with the help of the videos and you will have the possibility to try performing the steps once using the real parts.
Remember that your goal is to learn the task and you will be tested on performing it by yourself later on.

3.2 AR-group: Training the task using the AR-based training platform

You will now perform the task with the help of the AR. Remember that your goal is to learn the task and you will be tested on performing it by yourself later on.

3.3 Final instructions for all groups

For all groups – after the training

	Thank you for your participation so far. Please come back at XXX to undertake the real task. But before leaving, could you answer me a few questions about your experience with the training system.

C.2.3. Demographic Questionnaire and Results

Demographic Questionnaire

			STD	AR	STD	STD	AR	STD	AR	STD	AR	STD	STD	AR	AR	AR	STD	AR	STD	AR	STD	AR				
QUESTIONS			T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	AVG AR	AVG STD	VAR	DEV
DEMOGRAPHIC QUESTIONNAIRE	AGE	*Years*	34	27	50	35	37	33	34	28	29	33	34	31	30	33	36	35	33	43	33	36	37,22	38,778	26,063	5,1052
	SEX	*(M/F)*	M	M	M	M	M	M	M	M	M	M	M	M	M	M	F	M	M	M	M	M	/			
	Years of Experience	*Years*	4	3	20+	6	5	4	5	4	3	5	4	5	4	4	5	5	4	10+	4	4	3,78	4	0,65	0,8062
	Familiarity with computer-assisted training systems?	*1: none 5: assiduous use*	2	2	1	2	2	2	1	2	3	2	1	2	3	3	3	4	2	1	2	3	2,33	1,8889	0,6928	0,8324
	Familiarity with VR systems?	*1: none 5: assiduous use*	1	2	1	1	1	1	1	1	3	1	1	1	1	1	1	1	1	1	1	2	1,33	1	0,2647	0,5145
	Familiarity with AR systems?	*1: none 5: assiduous use*	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1	1	1	1	1	2	1,00	1,2222	0,2222	0,4714
	Familiarity with robotic systems	*1: none 5: assiduous use*	3	3	4	3	3	4	4	3	4	3	3	3	2	2	3	3	2	3	2	3	3,00	3,1111	0,4085	0,6391
	Familiarity with videogames	*1: none 5: assiduous use*	3	4	1	4	2	3	3	5	5	2	3	5	4	4	3	3	2	2	2	2	3,56	2,8889	1,3595	1,166
	Skill level in performing assembly/disassembly tasks	*1: beginner 5: expert*	3	4	5	3	5	4	4	3	4	4	5	4	3	3	4	4	4	5	4	3	4,00	3,8889	0,5261	0,7254

C.2.4. Exploitation Questionnaire and Results

Exploitation Questionnaire

	QUESTIONS		AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	
			T2	T5	T7	T9	T12	T13	T14	T16	T18	T20	AVG	DEV STD
System's help to learn	Rate the degree to which you completed the required task after the training on the system	1: not at all 5: wholly completed	3	4	4	4	3	3	5	4	5	5	4,0	0,8
	Do you think the training on the system quickly enhanced your skill level in the trained task?	(Yes/No)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
	Overall, I am satisfied with how I can perform the task trained by the system	1: strongly disagree 5: strongly agree	4	4	5	4	5	3	5	5	4	5	4,4	0,7
	I would recommend the system for training the task I experienced	1: strongly disagree 5: strongly agree	3	4	3	4	4	3	3	4	5	5	3,8	0,8
	I think the system is a valuable training means	1: strongly disagree 5: strongly agree	3	3	3	3	4	3	3	4	3	4	3,3	0,5
Ease of use of the Demonstrator	It took a long practice-time before I understood how the system works	1: strongly disagree 5: strongly agree	2	1	2	1	2	1	1	1	1	1	1,3	0,5
	I was completely concentrated more on the task activities rather than on how to use the system used during the training session	1: strongly disagree 5: strongly agree	5	4	5	5	4	4	5	5	5	4	4,6	0,5
	Rate the difficult you experienced in carrying out the assigned task on the system	1: very difficult 5: easy	5	5	5	5	5	5	5	5	5	5	5,0	0,0
	How efficiently did you complete the required task on the training system following the indications provided?	1: inefficiently 5: very efficiently	5	5	5	5	5	4	5	5	4	4	4,7	0,5
	Rate the level of clarity of the instructions provided during the training on the system?	1: very unclear 5: totally clear	5	5	5	5	5	5	5	5	5	4	4,9	0,3
	Whenever you make a mistake using the system, to which extent could you recover easily?	1: very difficult 5: easy	4	5	4	5	4	5	4	4	4	5	4,4	0,5
	Whenever you make a mistake using the system, to which extent could you recover quickly?	1: very difficult 5: easy	4	4	4	3	5	4	4	4	4	4	4,0	0,5
	Overall, I felt the experience with the system comfortable	1: strongly disagree 5: strongly agree	5	5	4	5	4	5	5	4	5	4	4,6	0,5
	Overall, the system is easy-to-use.	1: strongly disagree 5: strongly agree	5	5	5	4	5	5	5	4	5	5	4,8	0,4

C.2.5. Results of the Performance Measuring

Evaluation Test: Detailed Performance Time(s) and Errors

DETAILED MEASURES BY STEP

		STD	AR	STD	STD	AR	STD	AR	STD	AR	STD	STD	AR	AR	AR	STD	AR	STD	AR	STD	AR	STD	AR
STEP 1: Place the ROLLER into the BELT ROLLER		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	**Number of Errors**	**Number of Errors**
Types of errors / Use of aids																							
With Non-Solved errors	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						
STEP 2: Place the ROLLER+BELT ROLLER		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	**STD**	**AR**
Types of errors / Use of aids																							
With Non-Solved errors	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						
STEP 3: Tighten the NUT of the roller		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	**STD**	**AR**
Types of errors / Use of aids																							
With Non-Solved errors	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Comments	*Description*																						
																						0	**0**
PARTIAL TASK PERFORMANCE TIME		**85**	**79**	**73**	**54**	**59**	**52**	**47**	**43**	**42**	**108**	**59**	**44**	**39**	**35**	**35**	**36**	**30**	**18**	**50**	**50**	**58,90** / **23,81**	**44,90** / **16,11**

DETAILED MEASURES BY STEP

		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	STD	AR
STEP 4: Position the LOWER sensor at the support plate		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	**STD**	***AR***
Types of errors / Use of aids																							
With Non-Solved errors	*0/1*	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	5	1
With Errors solved with aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	2
With aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						
STEP 5: Tighten the screws fixing the lower sensor (1st screw)		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	**STD**	***AR***
Types of errors / Use of aids																							
With Non-Solved errors	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						
STEP 6: Tighten the screws fixing the lower sensor (2nd screw)		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	**STD**	***AR***
Types of errors / Use of aids																							
With Non-Solved errors	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						
STEP 7: Position the UPPER sensor at the support plate		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	**STD**	***AR***
Types of errors / Use of aids																							
With Non-Solved errors	*0/1*	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	5	1
With Errors solved with aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						
STEP 8: Tighten the screws holding the upper sensor (1st screw)		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	**STD**	***AR***
Types of errors / Use of aids																							
With Non-Solved errors	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						
STEP 9: Tighten the screws holding the upper sensor (2nd screw)		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	**STD**	***AR***
Types of errors / Use of aids	*0/1*																						
With Non-Solved errors	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						
																						10	**4**
PARTIAL TASK PERFORMANCE TIME		**410**	**205**	**251**	**145**	**174**	**213**	**183**	**154**	**120**	**169**	**149**	**205**	**224**	**149**	**168**	**264**	**121**	**122**	**130**	**178**	**191,00** **86,23**	**182,40** **44,92**

DETAILED MEASURES BY STEP

STEP 10: Position the support plate into the structure		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	STD	AR
Types of errors / Use of aids																							
With Non-Solved errors	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	0/1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0
With aids	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						

STEP 11: Tighten the fixing screws of the support plate (screws behind electronic board)		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	STD	AR
Types of errors / Use of aids																							
With Non-Solved errors	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	0/1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
With aids	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*	Wrong Screw																					

STEP 12: Tighten the fixing screws of the support plate (screws behind electronic board)		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	STD	AR
Types of errors / Use of aids																							
With Non-Solved errors	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	STD		AR	
																					3		0	
PARTIAL TASK PERFORMANCE TIME	**92**	**101**	**93**	**51**	**62**	**70**	**50**	**35**	**32**	**51**	**28**	**60**	**32**	**28**	**36**	**40**	**29**	**25**	**82**	**41**	**56,70**	**25,68**	**47,10**	**22,85**

DETAILED MEASURES BY STEP

STEP 13: Fix the electronic board into the support plate		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	STD	AR
Types of errors / Use of aids																							
With Non-Solved errors	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						

STEP 14: Connect the down sensor cable of the electronic board		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	STD	AR
Types of errors / Use of aids																							
With Non-Solved errors	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						

STEP 15: Connect the up sensor cable of the electronic board		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	STD	AR
Types of errors / Use of aids																							
With Non-Solved errors	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						

STEP 16: Keep the sensor cables tidy with nylon clip		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	STD	AR
Types of errors / Use of aids																							
With Non-Solved errors	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						

STEP 17: Connect the drive cable from the Electronic Board		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	STD	AR
Types of errors / Use of aids																							
With Non-Solved errors	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	0/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	STD		AR	
																					0		0	
PARTIAL TASK PERFORMANCE TIME	**60**	**49**	**50**	**14**	**104**	**45**	**45**	**42**	**37**	**32**	**60**	**35**	**35**	**19**	**28**	**35**	**15**	**20**	**38**	**31**	**38,40**	**16,39**	**41,00**	**24,03**

DETAILED MEASURES BY STEP

		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	STD	AR
STEP 18: Position the actuator cover		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	**STD**	***AR***
Types of errors / Use of aids																							
With Non-Solved errors	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						
STEP 19: Tighten the fixing screws of the actuator		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	**STD**	***AR***
Types of errors / Use of aids																							
With Non-Solved errors	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	*0/1*	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
With aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						
STEP 20: Tighten the fixing screws of the actuator		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	**STD**	***AR***
Types of errors / Use of aids																							
With Non-Solved errors	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						
STEP 21: Tighten the fixing screws of the actuator		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	**STD**	***AR***
Types of errors / Use of aids																							
With Non-Solved errors	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						
STEP 22: Tighten the fixing screws of the actuator		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	**STD**	***AR***
Types of errors / Use of aids																							
With Non-Solved errors	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						
																						0	**1**
PARTIAL TASK PERFORMANCE TIME		**130**	**187**	**155**	**96**	**34**	**193**	**155**	**130**	**103**	**115**	**108**	**216**	**130**	**99**	**93**	**105**	**60**	**70**	**93**	**98**	**117,30** / **37,24**	**119,70** / **54,12**

DETAILED MEASURES BY STEP

STEP 23: Position the clamp into the actuator stem		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	**STD**	***AR***
Types of errors / Use of aids																							
With Non-Solved errors	*0/1*	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	1	0	0	0	4	1
With Errors solved with aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						
STEP 24: Tighten the fixing screws of the clamp		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	**STD**	***AR***
Types of errors / Use of aids																							
With Non-Solved errors	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						
STEP 25: Tighten the fixing screws of the clamp		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	**STD**	***AR***
Types of errors / Use of aids																							
With Non-Solved errors	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With Errors solved with aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With errors solved without any aid	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
With aids	*0/1*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comments	*Description*																						
																						4	**1**
PARTIAL TASK PERFORMANCE TIME		**117**	**81**	**76**	**45**	**72**	**68**	**82**	**167**	**50**	**50**	**40**	**55**	**67**	**31**	**133**	**41**	**45**	**39**	**51**	**71**	**79,20** / **44,34**	**58,90** / **18,24**
																				TOTAL		**17**	**6**

Bibliography

[AAYS05] ALON J., ATHITSOS V., YUAN Q., SCLAROFF S.: Simultaneous localization and recognition of dynamic hand gestures. In *IEEE Workshop on Motion and Video Computing (WACV/MOTIONS '05)* (2005), vol. 2, pp. 254 –260. 77

[Ada71] ADAMS J. A.: A closed-loop theory of motor learning. *Journal of Motor Behavior 3*, 2 (1971), 111–149. 23

[And82] ANDERSON J. R.: Acquisition of cognitive skill. *Psychological Review 89*, 4 (1982), 369–406. 20, 84, 87, 88

[Ang02] ANG B. C.: The quest for nursing excellence. *Singapore Medical Journal 43*, 10 (2002), 493–495. 22

[APH*03] AGRAWALA M., PHAN D., HEISER J., HAYMAKER J., KLINGNER J., HANRAHAN P., TVERSKY B.: Designing effective step-by-step assembly instructions. *ACM Transactions on Graphics (TOG) 22*, 3 (2003), 828–837. 88

[AQP*01] ANTOLÍ A., QUESADA J. F., PSICOLOGÍA F. D., CAÑAS J. J., CAÑAS J. J.: The role of working memory on measuring mental models of physical systems. In *Psicológica* (2001), vol. 22, pp. 25–42. 88

[AV10] AKL A., VALAEE S.: Accelerometer-based gesture recognition via dynamic-time warping, affinity propagation, & compressive sensing. In *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP '10)* (2010), pp. 2270 –2273. 77, 110

[AZ97] ACCOT J., ZHAI S.: Beyond Fitts' law: models for trajectory-based HCI tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (1997), ACM, pp. 295–302. 20

[Azu97] AZUMA R. T.: A Survey of Augmented Reality. *Presence 6* (1997), 355–385. 35

[BA06] BURR D., ALAIS D.: Combining visual and auditory information. *Progress in Brain Research 155* (2006), 243–258. 56

[BAB*94] BERGAMASCO M., ALLOTTA B., BOSIO L., FERRETTI L., PARRINI G., PRISCO G., SALSEDO F., SARTINI G.: An arm exoskeleton system for teleoperation and virtual environments applications. In *Proceedings of the IEEE International Conference on Robotics and Automation* (1994), vol. 2, pp. 1449–1454. 52

[Ban86] BANDURA A.: *Social Foundations of Thought and Action: A Social Cognitive Theory*, 1st ed. Prentice Hall, 1986. 29

[Bat54] BATTIG W.: The effect of kinesthetic, verbal, and visual cues on the acquisition of a lever-positioning skill. *Journal of Experimental Psychology 47*, 5 (1954), 371–380. 56

[Bau97] BAUMANN R.: *Virtual Reality based Laparoscopic Surgery Training Environments*. PhD thesis, École Polytechnique Fédéral de Lausanne, 1997. 59

[BBb*05] BROWN D. G., BAILLOT Y., BAILEY M. P., PFLUGER K. C., MAASSEL P., THOMAS J., JULIER S. J.: Using Augmented Reality to enhance fire support team training. In *Proceedings of the Interservice/Industry Training, Simulation, & Education Conference (I/ITSEC '05)* (2005). 8

[BBF*01] BUCCINO G., BINKOFSKI F., FINK G. R., FADIGA L., FOGASSI L., GALLESE V., SEITZ R. J., ZILLES K., RIZZOLATTI G., FREUND H. J.: Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *The European Journal of Neuroscience 13*, 2 (2001), 400–404. 19

[BC03] BURDEA G., COIFFET P.: *Virtual Reality Technology*, 2nd ed. John Wiley & Sons, 2003. 49, 51

[BCK*05] BIMBER O., CORIAND F., KLEPPE A., BRUNS E., ZOLLMANN S., LANGLOTZ T.: Superimposing pictorial artwork with projected imagery. In *ACM SIGGRAPH 2005 Courses (SIGGRAPH '05)* (2005), ACM. 41

[BDC*10] BLUTEAU J., DUBOIS M., COQUILLART S., GENTAZ E., PAYAN Y.: Vibrotactile guidance for trajectory following in computer aided surgery. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '10)* (2010), pp. 2085–2088. 52, 59

[BDJW07] BEHR J., DÄHNE P., JUNG Y., WEBEL S.: Beyond the web browser - X3D and immersive VR. In *IEEE Virtual Reality 2007. VR Tutorial and Workshop Proceedings [CD-ROM] : IEEE Symposium on 3D User Interfaces* (2007), IEEE Computer Society u.a., p. 5 p. 106

[BDL*10] BARDY B., DELIGNIÈRES D., LAGARDE J., MOTTET D., ZELIC G.: An enactive approach to perception-action and skill acquisition in Virtual Reality environments. In *Proceedings of the 3rd International Conference on Applied Human Factors and Ergonomics (AHFE '10)* (2010), USA Publishing. 15, 16, 17, 18

[BDM99] BOLIA R. S., D'ANGELO W. R., MCKINLEY R. L.: Aurally aided visual search in three-dimensional space. *Human Factors 41*, 4 (1999), 664–669. 56

[BDR04] BEHR J., DÄHNE P., ROTH M.: Utilizing X3D for immersive environments. In *Proceedings of the 9th International Conference on 3D Web Technology (Web3D '04)* (2004), ACM, pp. 71–78. 106

[Ben04] BENNER P.: Using the Dreyfus Model of skill acquisition to describe and interpret skill acquisition and clinical judgment in nursing practice and education. *Bulletin of Science, Technology & Society 24*, 3 (2004), 188 –199. 22

[Ber67] BERNSTEIN N. A.: *The co-ordination and regulation of movements*. Pergamon, 1967. 16

[BFA07] BERGAMASCO M., FRISOLI A., AVIZZANO C.: Exoskeletons as man-machine interface systems for teleoperation and interaction in virtual environments. In *Advances in Telerobotics*, vol. 31 of *Springer Tracts in Advanced Robotics*. Springer Berlin / Heidelberg, 2007, pp. 61–76. 52

[BGW*02] BIMBER O., GATESY S. M., WITMER L. M., RASKAR R., ENCARNAÇÃO L. M.: Merging fossil specimens with computer-generated information. *Computer 35*, 9 (2002), 25–30. 41

[BHKN09] BLUM T., HEINING S. M., KUTTER O., NAVAB N.: Advanced training methods using an Augmented Reality ultrasound simulator. In *Proceedings of the 8th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '09)* (2009), IEEE Computer Society, pp. 177–178. 9

[BK04] BISCHOFF R., KAZI A.: Perspectives on Augmented Reality based human-robot interaction with industrial robots. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2004), vol. 4, pp. 3226–3231. 9

[BPG*06] BURKE J. L., PREWETT M. S., GRAY A. A., YANG L., STILSON F. R. B., COOVERT M. D., ELLIOT L. R., REDDEN E.: Comparing the effects of visual-auditory and visual-tactile feedback on user performance: a meta-analysis. In *Proceedings of the 8th International Conference on Multimodal Interfaces (ICMI '06)* (2006), ACM, pp. 108–117. 59

[BR05] BIMBER O., RASKAR R.: *Spatial Augmented Reality: Merging real and virtual worlds*. AK Peters Ltd, 2005. 37, 38, 40

[CB90] CAROLL W., BANDURA A.: Representational guidance of action production in observational guidance of action production in observational learning: A causal analysis. *Journal of Motor Behavior 22* (1990), 85–97. 29

[CBC03] CAKMAKCI O., BÈRARD F., COUTAZ J.: An Augmented Reality based learning assistant for electric bass guitar. In *In Proceedings of the 10th International Conference on Human-Computer Interaction (HCI '03)* (2003). 12

[CEG*09] CASADO S., ENGELKE T., GAVISH N., GUTIÉRREZ T., RODRÍGUEZ J., SÁNCHEZ E. J., TECCHIA F., WEBEL S.: A training system for skills transfer involved in industrial maintenance and assembly tasks. In *Enaction on SKILLS: International Conference on Multimodal Interfaces for Skills Transfer (SKILLS '09)* (2009), pp. 129–134. 59

[CMN83] CARD S. K., MORAN T. P., NEWELL A.: *The psychology of human-computer interaction*. CRC, 1983. 31

[Coh84] COHEN N. J.: Preserved learning capacity in amnesia: Evidence for multiple memory systems. *Neuropsychology of memory* (1984), 83–103. 88

[CS96] CARLSON R. A., SHIN J. C.: Practical schedules and subgoals instantiation in cascaded problem solving. *Journal of Experimental Psychology: Learning, Memory and Cognition 22*, 1 (1996), 157–168. 30

[CT04] CALVERT G. A., THESEN T.: Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris 98*, 1-3 (2004), 191–205. 34

[Däh08] DÄHNE P.: *Entwurf eines Rahmensystems für mobile Augmented-Reality-Anwendungen*. PhD thesis, TU Darmstadt, 2008. 108

[DBB08] DAVIDS K. W., BUTTON C., BENNETT S. J.: *Dynamics of skill acquisition : a constraints-led approach*. Human Kinetics, Champaign, Illinois, 2008. 21, 24

[DC04] DIEDERICH A., COLONIUS H.: Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics 66*, 8 (2004), 1388–1404. 34

[DCY10] DING Z., CHEN I., YEO S.: The development of a real-time wearable motion replication platform with spatial sensing and tactile feedback. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '10)* (2010), pp. 3919 –3924. 52

[DD80] DREYFUS S. E., DREYFUS H. L.: *A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition*. Tech. rep., University of California, Berkeley, 1980. 4, 22

[Deb98] DEBEVEC P.: Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '98)* (1998), ACM, pp. 189–198. 50

[Deb06] DEBEVEC P.: Image-based lighting. In *ACM SIGGRAPH 2006 Courses (SIGGRAPH '06)* (2006), ACM. 50

[DFS89] DONCHIN E., FABIANI M., SANDERS A.: The learning strategies program. An examination of the strategies in skill acquisition. *Acta Psychologica 71*, 1-3 (1989), 1–309. 26

[Don69] DONDERS F. C.: On the speed of mental processes. *Acta Psychologica 30* (1969), 412–431. 31

[dS97] DE'SPERATI C., STUCCHI N.: Recognizing the motion of a graspable object is guided by handedness. *NeuroReport 8*, 12 (1997), 2761–2765. 19

[DS01] DOYLE M. C., SNOWDEN R. J.: Identification of visual stimuli is improved by accompanying auditory stimuli: The role of eye movements and sound location. *Perception 30*, 7 (2001), 795–810. 33

[EB04] ERNST M. O., BÜLTHOFF H. H.: Merging the senses into a robust percept. *Trends in Cognitive Sciences 8*, 4 (2004), 162–169. 33

[EBW97] ELLIS S. R., BEGAULT D., WENZEL E. M.: Virtual environments as human-computer interfaces. In *Handbook of human-computer interaction*. Elsevier, 1997, ch. 8. 56

[ECR09] ELLIOTT L. R., COOVERT M. D., REDDEN E. S.: Overview of meta-analyses investigating vibrotactile versus visual display options. In *Proceedings of the 13th International Conference on Human-Computer Interaction. Part II: Novel Interaction Methods and Techniques* (2009), Springer Berlin / Heidelberg, pp. 435–443. 59

[EvERD10] ELLIOTT L., VAN ERP J., REDDEN E., DUISTERMAAT M.: Field-based validation of a tactile navigation device. *IEEE Transactions on Haptics 3*, 2 (2010), 78–87. 59

[FBG*89] FABIANI M., BUCKLEY J., GRATTON G., COLES M., DONCHIN E., LOGIE R.: The training protocol of complex task performance. *Acta Psychologica 71* (1989), 259–299. 27, 30

[FCAB02] FORSTER B., CAVINA-PRATESI C., AGLIOTI S. M., BERLUCCHI G.: Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Experimental Brain Research 143*, 4 (2002), 480–487. 34

[Fer91] FERRIN F. J.: Survey of helmet tracking technologies. *Large Screen Projection, Avionic, and Helmet-Mounted Displays 1456*, 1 (1991), 86–94. 45

[FHB03] FÖRSTER J., HIGGINS E. T., BIANCO A. T.: Speed/accuracy decisions in task performance: Built-in trade-off or separate strategic concerns? *Organizational Behavior and Human Decision Processes 90*, 1 (2003), 148–164. 30, 31

[Fia05] FIALA M.: ARTag, a fiducial marker system using digital techniques. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)* (2005), vol. 2, pp. 590–596. 43

[Fia10] FIALA M.: Designing highly reliable fiducial markers. *IEEE Transactions on Pattern Analysis and Machine Intelligence 32* (2010), 1317–1324. 43

[Fit54] FITTS P. M.: The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychololgy 47*, 6 (1954), 381–391. 4, 20, 31

[Fli10] FLINT S.: Opportunity for investment and development Augmented Reality for improved sports performance, 2010. 12

[FMHS93] FEINER S., MACINTYRE B., HAUPT M., SOLOMON E.: Windows on the world: 2D windows for 3D Augmented Reality. In *Proceedings of the 6th Annual ACM symposium on User Interface Software and Technology (UIST '93)* (1993), ACM, pp. 145–155. 40

[FMS93] FEINER S., MACINTYRE B., SELIGMANN D.: Knowledge-based Augmented Reality. *Commun. ACM 36*, 7 (1993), 53–62. 7

[FN03] FOXLIN E., NAIMARK L.: Vis-tracker: A wearable vision-inertial self-tracker. In *Proceedings of the IEEE Conference on Virtual Reality (VR '03)* (2003), pp. 199–206. 43

[FOO93] FEIN R. M., OLSON G. M., OLSON J. S.: A mental model can help with learning to operate a complex device. In *CHI '93: INTERACT '93 and CHI '93 Conference Companion on Human Factors in Computing Systems* (1993), ACM, pp. 157–158. 88

[Fox02] FOXLIN E.: Motion tracking requirements and technologies. *Handbook of Virtual Environments: Design, Implementation and Applications* (2002), 163–210. 46

[Fra06] FRANKLIN M.: The lessons learned in the application of Augmented Reality. In *Virtual Media for Military Applications* (2006), pp. 30/1–30/8. 9

[Fra08] FRAUNHOFER INSTITUTE FOR COMPUTER GRAPHICS RESEARCH IGD: instantreality, 2008. http://www.instantreality.org/; website checked on May 2011. 63, 106

[FW89] FREDERIKSEN J. R., WHITE B. Y.: An approach to training based upon principled task decomposition. *Acta Psychologica 71*, 1-3 (1989), 89–146. 27, 30

[GB03] GREENE J. O., BURLESON B. R.: *Handbook of communication and social interaction skills*. Routledge, 2003. 20

[GBK*08] GOPHER D., BARDY B., KRUPENIA S., GAVISH N., KRUPENIA V., KALLAI A., LEV D., HOFFMANN P.: Assessment of potential benefits and pitfalls of multimodal virtual training (D.RA1.5). In *SKILLS, reporting and documentation project Series*. 2008. 33

[GBL04] GUIARD Y., BEAUDOUIN-LAFON M.: Fitts' law 50 years later: applications and contributions from human-computer interaction. *International Journal of Human-Computer Studies - Special issue 61*, 6 (2004), 747–750. 31

[GC02] GUPTA P., COHEN N. J.: Theoretical and computational analysis of skill learning, repetition priming, and procedural memory,. *Psychological Review 109*, 2 (2002), 401–448. 84, 88

[GD02] GRÈZES J., DECETY J.: Does visual perception of object afford action? Evidence from a neuroimaging study. *Neuropsychologia 40*, 2 (2002), 212–222. 19

[GFM*03] GAUSEMEIER J., FRUEND J., MATYSCZOK C., BRUEDERLIN B., BEIER D.: Development of a real time image based object recognition method for mobile AR-devices. In *Proceedings of the 2nd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa (AFRIGRAPH '03)* (2003), ACM, pp. 133–139. 40

[GKG10] GOPHER D., KRUPENIA S., GAVISH N.: Skill training in multimodal virtual environments. In *Proceedings of the 3rd International Conference on Applied Human Factors and Ergonomics (AHFE '10)* (2010), USA Publishing. 60, 61

[GKRS01] GEIGER C., KLEINNJOHANN B., REIMANN C., STICHLING D.: Mobile AR4ALL. In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality (ISAR '00)* (2001), pp. 181 –182. 40

[GMBT09] GOSSELIN F., MǴARD C., BOUCHIGNY S., TAHA F.: A training system for the transfer of skills involved in maxillo facial surgery. In *Enaction on SKILLS: International Conference on Multimodal Interfaces for Skills Transfer (SKILLS '09)* (2009), pp. 143–

148. 10

[Gop93] GOPHER D.: The skill of attention control: Acquisition and execution of attention strategies. *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (1993), 299–322. 26, 29

[Gop04] GOPHER D.: Control processes in the formation of task units. In *28th International Congress of Psychology, Bejing, China* (2004). 15

[Gop07] GOPHER D.: *Emphasis change as a training protocol for high-demands tasks*. Oxford University Press, 2007. 29, 30, 60, 65

[GWB94] GOPHER D., WEIL M., BAREKET T.: Transfer of a skill from a computer game trainer to flight. *Human Factors 36*, 3 (1994), 387–405. 29

[GWS89] GOPHER D., WEIL M., SIEGEL D.: Practice under changing priorities: An approach to the training of complex skills. *Acta Psychologica 71*, 1-3 (1989), 147–177. 26

[HA09] HAHNE U., ALEXA M.: Depth imaging by combining time-of-flight and on-demand stereo. In *Proceedings of the DAGM Workshop on Dynamic 3D Imaging (Dyn3D '09)* (2009), Springer Berlin / Heidelberg, pp. 70–83. 51

[HB69] HANDEL S., BUFFARDI L.: Using several modalities to perceive one temporal pattern. *Quarterly Journal of Experimental Psychology 21*, 3 (1969), 256–266. 33

[HBO05] HENRYSSON A., BILLINGHURST M., OLLILA M.: Face to face collaborative AR on mobile phones. In *Proceedings of the 4th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '05)* (2005), IEEE Computer Society, pp. 80–89. 40

[HDBB97] HANDFORD C., DAVIDS K., BENNETT S., BUTTON C.: Skill acquisition in sport: Some applications of an evolving practice ecology. *Journal of Sports Sciences 15*, 6 (1997), 621–640. 28

[Hec07] HECHT B.: Wikeye - using magic lenses to explore georeferenced wikipedia content. In *Proceedings of the 3rd International Workshop on Pervasive Mobile Interaction Devices (PERMID '07)* (2007). 40

[Hel82] HELLER M. A.: Visual and tactual texture perception: intersensory cooperation. *Perception & Psychophysics 31*, 4 (1982), 339–344. 34

[HKT*10] HAGN U., KONIETSCHKE R., TOBERGTE A., NICKL M., JÖRG S., KÜBLER B., PASSIG G., GRÖGER M., FRÖHLICH F., SEIBOLD U., LE-TIEN L., ALBU-SCHÄFFER A., NOTHHELFER A., HACKER F., GREBENSTEIN M., HIRZINGER G.: DLR MiroSurge: a versatile system for research in endoscopic telesurgery. *International Journal of Computer Assisted Radiology and Surgery 5* (2010), 183–193. 10.1007/s11548-009-0372-4. 10

[HNJ*08] HAGN U., NICKL M., JÖRG S., PASSIG G., BAHLS T., NOTHHELFER A., HACKER F., LE-TIEN L., ALBU-SCHÄFFER A., KONIETSCHKE R., GREBENSTEIN M., WARPUP R., HASLINGER R., FROMMBERGER M., HIRZINGER G.: The DLR MIRO: a versatile lightweight robot for surgical applications. *Industrial Robot: An International Journal, Special Issue: Theme Title: Medical Robots 35*, 4 (2008), 324–336. 10

[HOK*08] HAGN U., ORTMAIER T., KONIETSCHKE R., KUBLER B., SEIBOLD U., TOBERGTE A., NICKL M., JÖRG S., HIRZINGER G.: Telemanipulator for remote minimally invasive surgery. *IEEE Robotics Automation Magazine 15*, 4 (2008), 28 –38. 10

[HOMS06] HOSHI T., OKADA A., MAKINO Y., SHINODA H.: A whole body artificial skin based on cell-bridge networking system. In *Proceedings of the 3rd International Conference on Networked Sensing Systems (INSS'06)* (2006), pp. 55–60. 48

[HP91] HERRNSTEIN R. J., PRELEC D.: Melioration: A theory of distributed choice. *Journal of Economic Perspectives 5*, 3 (1991), 137–56. 26, 29

[HSA*08] HULIN T., SAGARDIA M., ARTIGAS J., SCHÄTZLE S., KREMER P., PREUSCHE C.: Human-scale bimanual haptic interface. In *Enactive '08* (2008). 59

[IMP*05] ISSENBERG S. B., MCGAGHIE W. C., PETRUSA E. R., GORDON D. L., SCALESE R. J.: Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical Teacher 27*, 1 (2005), 10–28. 10

[JEW*97] JEBARA T., EYSTER C., WEAVER J., STARNER T., PENTLAND A.: Stochasticks: augmenting the billiards experience with probabilistic vision and wearable computers. In *Proceedings of the 1st International Symposium on Wearable Computers* (1997), pp. 138–145. 12

[JFDB07] JUNG Y., FRANKE T., DÄHNE P., BEHR J.: Enhancing X3D for advanced MR appliances. In *Proceedings of the 12th International Conference on 3D Web Technology (Web3D '07)* (2007), ACM, pp. 27–36. 106

[JWO*10] JUNG Y., WEBEL S., OLBRICH M., DREVENSEK T., FRANKE T., ROTH M., FELLNER D.: Interactive textures as spatial user interfaces in X3D. In *Proceedings of the 15th International Conference on Web 3D Technology (Web3D '10)* (2010), ACM, pp. 147–150. 122

[KB99] KATO H., BILLINGHURST M.: Marker tracking and HMD calibration for a video-based Augmented Reality conferencing system. In *Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality* (1999), IEEE Computer Society, pp. 85–. 8, 12, 42

[KD04] KLEIN G., DRUMMOND T.: Sensor fusion and occlusion refinement for tablet-based AR. In *Proceedings of the 3rd IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '04)* (2004), pp. 38–47. 45

[Kee68] KEELE S. W.: Movement control in skilled motor performance. *Psychological Bulletin 70* (1968), 387–403. 23

[KG02] KUO B. C., GOLNARAGHI F.: *Automatic Control Systems*, 8th ed. Wiley, 2002. 23

[KHYN02] KURIHARA K., HOSHINO S., YAMANE K., NAKAMURA Y.: Optical motion capture system with pan-tilt camera tracking and real time data processing. In *Proceedings of the*

IEEE International Conference on Robotics and Automation (ICRA'02) (2002), vol. 2, pp. 1241 –1248. 46

[Kie88] KIERAS D.: What mental model should be taught: Choosing instructional content for complex engineered systems. In *Intelligent tutoring systems: Lessons learned* (1988), Lawrence Erlbaum Associates, pp. 85–111. 88

[KKCL05] KE C., KANG B., CHEN D., LI X.: An Augmented Reality-based application for equipment maintenance. In *ACII* (2005), pp. 836–841. 8

[KM07] KLEIN G., MURRAY D.: Parallel tracking and mapping for small ar workspaces. In *Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '07)* (2007), IEEE Computer Society, pp. 1–10. 44

[KM09] KLEIN G., MURRAY D.: Parallel tracking and mapping on a camera phone. In *Proceedings of the 8th IEEE International Symposium on Mixed and Augmented Reality (ISMAR '09)* (2009), IEEE Computer Society, pp. 83–86. 44

[KPKL07] KIM J., PARK J., KIM H., LEE C.: HCI (human computer interaction) using multi-touch tabletop display. *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing* (2007), 391–394. 110

[LA08] LIVINGSTON M. A., AI Z.: The effect of registration error on tracking distant augmented objects. In *Proceedings of the 7th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '08)* (2008), IEEE Computer Society, pp. 77–86. 71

[LB07] LIEBERMAN J., BREAZEAL C.: TIKL: Development of a wearable vibrotactile feedback suit for improved human motor learning. *IEEE Transactions on Robotics 23*, 5 (2007), 919 –926. 52

[Lew04] LEWIS J. R.: Cover story: in the eye of the beholder. *IEEE Spectrum 41*, 5 (2004), 24–28. 37

[LFA*09] LI S., FRISOLI A., AVIZZANO C. A., RUFFALDI E., BERGAMASCO M.: Practical evaluation of bimanual haptic desktop for upper-limb stroke rehabilitation. In *Enaction on SKILLS: International Conference on Multimodal Interfaces for Skills Transfer (SKILLS '09)* (2009), pp. 97–104. 11

[LL04] LALANNE C., LORENCEAU J.: Crossmodal integration for perception and action. *Journal of Physiology-Paris 98*, 1-3 (2004), 265–279. 34

[LMHK02] LEDERMAN S. J., MORGAN T., HAMILTON C., KLATZKY R. L.: Integrating multimodal information about surface texture via a probe: Relative contributions of haptic and touch-produced sound sources. In *Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS '02)* (2002), IEEE Computer Society, pp. 97–104. 34

[Log88] LOGAN G. D.: Toward an instance theory of automatization. *Psychological Review 95*, 4 (1988), 492–527. 32

[LZA*09] LAGARDE J., ZELIC G., AVIZZANO C., LIPPI V., RUFFALDI E., ZALMANOV H., EREV-YEHENE V., GILLIAN N., O'MODRAIN S., MOTTET D., GOPHER D.: The light weight juggling training system for juggling. In *Enaction on SKILLS: International Conference on Multimodal Interfaces for Skills Transfer (SKILLS '09)* (2009), pp. 91–95. 11

[Mac92] MACKENZIE I. S.: Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction 7*, 1 (1992), 91–139. 31, 32

[Mag07] MAGILL R. A.: *Motor learning and control: Concepts and applications*. McGraw-Hill, 2007. 28

[Mal08] MALERCZYK C.: Gestural interaction using feature classification. In *Proceedings of the 5th International Conference on Articulated Motion and Deformable Objects (AMDO '08)* (2008), Springer Berlin / Heidelberg, pp. 228–237. 45

[May01] MAYER R. E.: *Multimedia Learning*. Cambridge University Press, 2001. 55

[May02] MAYER R.: Aids to computer-based multimedia learning. *Learning and Instruction 12*, 1 (2002), 107–119. 70

[MBS05] MANTEL B., BARDY B., STOFFREGEN T.: Intermodal specification of egocentric distance in a target reaching task. In *Studies in Perception and Action VIII* (2005), pp. 173–176. 33

[Mic10] MICROSOFT: Kinect, 2010. http://www.xbox.com/en-US/kinect; website checked on May 2011. 51

[MJKM04] MILLER N., JENKINS O., KALLMANN M., MATARIC M.: Motion capture from inertial sensing for untethered humanoid teleoperation. In *Proceedings of the 4th IEEE/RAS International Conference on Humanoid Robots* (2004), vol. 2, pp. 547–565. 47

[MK94] MILGRAM P., KISHINO F.: A taxonomy of Mixed Reality visual displays. *IEICE Transactions on Information Systems E77-D*, 12 (1994). 35

[MLB04] MÖHRING M., LESSIG C., BIMBER O.: Video see-through AR on consumer cell-phones. In *Proceedings of the 3rd IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '04)* (2004), pp. 252–253. 40

[MMM*05] MURRAY M. M., MOLHOLM S., MICHEL C. M., HESLENFELD D. J., RITTER W., JAVITT D. C., SCHROEDER C. E., FOXE J. J.: Grabbing your ear: Rapid Auditory-Somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cerebral Cortex 15*, 7 (2005), 963 –974. 34

[MSB*04] MAUL H., SCHARF A., BAIER P., WUSTEMANN M., GUNTER H. H., GEBAUER G., SOHN C.: Ultrasound simulators: experience with the SonoTrainer and comparative review of other training systems. *Ultrasound in Obstetrics and Gynecology* (2004), 581–585. 10

[MTB*07] MORRIS D., TAN H., BARBAGLI F., CHANG T., SALISBURY K.: Haptic feedback enhances force skill learning. In *Proceedings of the 2nd Joint EuroHaptics Conference* *and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (World Haptics '07)* (2007), IEEE Computer Society, pp. 21–26. 56

[New91] NEWELL K.: Motor skill acquisition. *Annual Review of Psychology 42* (1991), 213–237. 28

[NF02] NAIMARK L., FOXLIN E.: Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker. In *Proceedings of the 1st IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '02)* (2002), pp. 27–36. 43

[NM00] NOVICK L., MORSE D.: Folding a fish, making a mushroom: The role of diagrams in executing assembly procedures. *Memory & Cognition 28*, 7 (2000), 1242–1256. 88

[Nok11] NOKIA: Qt – cross-platform application and UI framework, 2011. http://qt.nokia.com/; website checked on May 2011. 122

[Nor87] NORMAN D. A.: Some observations on mental models. In *Human-computer interaction*. Morgan Kaufmann Publishers Inc., 1987, pp. 241–244. 88

[NR93] NEWELL A., ROSENBLOOM P. S.: Mechanisms of skill acquisition and the law of practice. In *The Soar papers (vol. 1)*. MIT Press, 1993, pp. 81–135. 20, 32

[OGL08] OLWAL A., GUSTAFSSON J., LINDFORS C.: Spatial Augmented Reality on industrial CNC-machines. In *Proceedings of the SPIE: The Engineering Reality of Virtual Reality* (2008), vol. 6804, pp. 680409.1–680409.9. 8

[Ope03] OPENSCENEGRAPH: OpenSceneGraph documentation, 2003. http://www.openscenegraph.org; website checked on May 2011. 104

[Org11] ORGANIC MOTION INC.: Organic Motion, 2011. http://www.organicmotion.com/; website checked on May 2011. 47

[OVB04] OVING A. B., VELTMAN J. A., BRONKHORST A. W.: Effectiveness of 3-D audio for warnings in the cockpit. *The International Journal of Aviation Psychology 14*, 3 (2004), 257–276. 56

[PCS*00] PATRICK E., COSGROVE D., SLAVKOVIC A., RODE J. A., VERRATTI T., CHISELKO G.: Using a large projection screen as an alternative to head-mounted displays for virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'00)* (2000), ACM, pp. 478–485. 38

[PFV98] PRYOR H. L., FURNESS T. A., VIIRRE E.: The virtual retinal display: A new display technology using scanned laser light. *Human Factors and Ergonomics Society Annual Meeting Proceedings 42* (1998), 1570–1574. 37

[PKT*09] PREUSCHE C., KONIETSCHKE R., TOBERGTE A., RUFFALDI E., WEBEL S., BOCKHOLT U.: Development of a multimodal skills trainer for minimally invasive telerobotic surgery. In *Enaction on SKILLS: International Conference on Multimodal Interfaces for Skills Transfer (SKILLS '09)* (2009), pp. 153–160. 10

[PMGD87] PROTEAU L., MARTENIUK R. G., GIROUARD Y., DUGAS C.: On the type of information used to control and learn an aiming movement after moderate and extensive training. *Human Movement Science 6*, 2 (1987), 181 – 199. 56

[PMS95] PETERSON C., MAIER S. F., SELIGMAN M. E. P.: *Learned Helplessness: A Theory for the Age of Personal Control*. Oxford University Press, USA, 1995. 25, 26

[Pop01] POPESCU G. V.: *Design and performance analysis of a Virtual Reality-based telerehabilitation system*. PhD thesis, Rutgers University New Brunswik (NJ, USA), 2001. 51

[PR06] POLYS N. F., RAY A.: Supporting Mixed Reality interfaces through X3D specification. In *Workshop Proceedings of the IEEE Conference on Virtual Reality (VR '06): Specification of Mixed Reality User Interfaces* (2006), pp. 29–32. 104

[Pro06] PROFFIT D.: Distance perception. *Current Directions in Psychological Science 15*, 3 (2006), 131–135. 19

[PRSGA*10] PORTILLO-RODRIGUEZ O., SANDOVAL-GONZALEZ O. O., AVIZZANO C., RUFFALDI E., BERGAMASCO M.: Capturing and training motor skills. *Human-Robot Interaction* (2010). 15

[PRWEH98] PHILBIN D., RIBARSKY W., WALKER N., ELLIS HUBBARD C.: Training in virtual environments: analysis of task appropriateness. In *Proceedings of the IEEE Virtual Reality Annual International Symposium* (1998), p. 210. 83

[PS09] PETERSEN N., STRICKER D.: Fast hand detection using posture invariant constraints. In *Proceedings of the 32nd Annual German Conference on Advances in Artificial Intelligence (KI '09)* (2009), Springer-Verlag, pp. 106–113. 45

[PSS*04] PARKER S. P. A., SMITH S. E., STEPHAN K. L., MARTIN R. L., MCANALLY K. I.: Effects of supplementing head-down displays with 3-D audio during visual target acquisition. *The International Journal of Aviation Psychology 14*, 3 (2004), 277–295. 56

[Rab89] RABINER L. R.: A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE 77*, 2 (1989), 257–286. 77, 110

[RCC93] ROYER J. M., CISERO C. A., CARLO M. S.: Techniques and procedures for assessing cognitive skills. *Review of Educational Research 63*, 2 (1993), 201. 31

[RD05] ROSTEN E., DRUMMOND T.: Fusing points and lines for high performance tracking. In *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05)* (2005), vol. 2, pp. 1508–1515. 43

[Rek95] REKIMOTO J.: The magnifying glass approach to Augmented Reality systems. In *Proceedings of the International Conference on Artificial Reality and Tele-Existence Conference on Virtual Reality Software and Technology (ICAT/VRST '95)* (1995), pp. 123–132. 39

[RFF*09] RUFFALDI E., FILIPPESCHI A., FRISOLI A., AVIZZANO C., BARDY B., GOPHER D., BERGAMASCO M.: SPRINT: a training system for skills transfer in rowing. In *Enac-*

tion on SKILLS: International Conference on Multimodal Interfaces for Skills Transfer (SKILLS '09) (2009), pp. 87–90. 11

[RGH*10] RHIENMORA P., GAJANANAN K., HADDAWY P., SUEBNUKARN S., DAILEY M. N., SUPATARATARN E., SHRESTHA P.: Haptic Augmented Reality dental trainer with automatic performance assessment. In *Proceeding of the 14th International Conference on Intelligent User Interfaces (IUI '10)* (2010), ACM, pp. 425–426. 10

[RH94] ROHLF J., HELMAN J.: IRIS Performer: a high performance multiprocessing toolkit for real-time 3D graphics. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '94)* (1994), ACM, pp. 381–394. 104

[RHF94] ROLLAND J. P., HOLLOWAY R. L., FUCHS H.: A comparison of optical and video see-through head-mounted displays. In *Proceedings of SPIE: Telemanipulator and Telepresence Technologies* (1994), pp. 293–307. 38, 40

[RMW09] ROBERTSON C. M., MACINTYRE B., WALKER B. N.: An evaluation of graphical context as a means for ameliorating the effects of registration error. *IEEE Transactions on Visualization and Computer Graphics 15* (2009), 179–192. 71

[RSKM98] REINERS D., STRICKER D., KLINKER G., MÜLLER S.: Augmented Reality for construction tasks: Doorlock assembly. In *Proceedings of the 1st IEEE and ACM International Workshop on Augmented Reality (IWAR '98)* (1998), AK Peters, pp. 31–46. 8

[RVB02] REINERS D., VOSS G., BEHR J.: A multi-thread safe foundation for scene graphs and its extension to clusters. *Proceedings of the Eurographics Workshop on Parallel Graphics and Visualisation* (2002). 104, 106, 109

[RWLB01] RASKAR R., WELCH G., LOW K.-L., BANDYOPADHYAY D.: Shader lamps: Animating real objects with image-based illumination. In *Proceedings of the 12th Eurographics Workshop on Rendering Techniques* (2001), Springer Berlin / Heidelberg, pp. 89–102. 41

[SC92] STRAUSS P. S., CAREY R.: An object-oriented 3D graphics toolkit. In *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '02)* (1992), ACM, pp. 341–349. 105

[Sch75] SCHMIDT R.: A schema theory of discrete motor skill learning. *Psychological Review 82*, 4 (1975), 225–260. 23

[SCHB01] STETTEN G., CHIB V., HILDEBRAND D., BURSEE J.: Real time tomographic reflection: phantoms for calibration and biopsy. In *Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR '01)* (2001), IEEE Press, pp. 11 –19. 39

[SD05] STOLL J., DUPONT P.: Passive markers for ultrasound tracking of surgical instruments. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2005*, vol. 3750/2005 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2005, pp. 41–48. 42

[SEWP10] SCHÄTZLE S., ENDE T., WÜSTHOFF T., PREUSCHE C.: VibroTac: An ergonomic and versatile usable vibrotactile feedback device. In *Proceedings of the 9th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN '10)* (2010), IEEE Computer Society. 51, 52, 59, 86

[SGR98] SHEBILSKE W., GOETTL B., REIGAN J.: Individual and team protocols for training complex skills in laboratory and applied settings. *Attention and Performance, XVII: Cognitive Regulation of Performance Theory & Applications: Interaction of Theory and Application* (1998), 401–432. 29

[SGTRF09] SANDOVAL-GONZALEZ O., TRIPICCHIO P., RUFFALDI E., FILIPPESCHI A.: Skills accelerators using multimodal systems for boxing training. In *Enaction on SKILLS: International Conference on Multimodal Interfaces for Skills Transfer (SKILLS '09)* (2009), pp. 79–85. 11

[SHPH06] SCHÄTZLE S., HULIN T., PREUSCHE C., HIRZINGER G.: Evaluation of vibrotactile feedback to the human arm. In *Proceedings of the EuroHaptics* (2006), pp. 557–560. 51

[SIFW03] SUDDERTH E. B., IHLER A. T., FREEMAN W. T., WILLSKY A. S.: Nonparametric belief propagation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03) 1* (2003), 605–612. 45

[SII*10] SUDDERTH E. B., IHLER A. T., ISARD M., FREEMAN W. T., WILLSKY A. S.: Nonparametric belief propagation. *Communications of the ACM 53*, 10 (2010), 95–103. 45

[Sin80] SINGER R. N.: *Motor Learning and Human Performance: An Application to Motor and Movement Behaviors*, 3rd ed. MacMillan (New York), 1980. 15

[Ski06] SKINNER B. F.: *Behavior of Organisms*. Copley Publishing Group, 2006. 28

[SKI07] SKILLS CONSORTIUM: Literature review on the definition of skills in the context of multimodal interfaces (D.RA1.1). In *SKILLS, reporting and documentation project Series*. 2007. 15, 19, 25, 26, 27, 28, 29, 30, 31, 32

[SKI09] SKILLS CONSORTIUM: Demonstration of advanced capturing prototypes (a- technology level, b- platform level, c- skills level) (D.RA3.5). In *SKILLS, reporting and documentation project Series*. 2009. 46, 48

[SKI11] SKILLS CONSORTIUM: SKILLS, 2006-2011. http://www.skills-ip.eu/; website checked on January 2012. 61

[SLSG03] SCHWALD B., LAVAL B. D., SA T. O., GUYNEMER R.: An Augmented Reality system for training and assistance to maintenance in the industrial context. In *Proceedings of the 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG '03)* (2003), pp. 425–432. 7

[SM04] SOUKOREFF R., MACKENZIE I.: Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies 61*, 6 (2004), 751–789. 20

[SMFW04] SUDDERTH E., MANDEL M., FREEMAN W., WILLSKY A.: Visual hand tracking using nonparametric belief propagation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop (CVPRW '04)* (2004), vol. 12, pp. 189–197. 45

[SP54] SUMBY W. H., POLLACK I.: Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America 26* (1954), 212–215. 33

[SPD04] SPENCE C., PAVANI F., DRIVER J.: Spatial constraints on visual-tactile cross-modal distractor congruency effects. *Cognitive, Affective, & Behavioral Neuroscience 4*, 2 (2004), 148–169. 34, 56

[SS77] SCHNEIDER, SHIFFRIN R. M.: Controlled and automatic human information processing: 1. detection, search, and attention. *Psychological Review 84*, 1 (1977), 1–66. 26

[SS06] SUPAN P., STUPPACHER I.: Interactive image based lighting in Augmented Reality. *Central European Seminar on Computer Graphics* (2006). 50

[SYSS89] SCHMIDT R. A., YOUNG D. E., SWINNEN S., SHAPIRO D. C.: Summary knowledge of results for skill acquisition: support for the guidance hypothesis. *Journal of Experimental Psychology. Learning, Memory, and Cognition 15*, 2 (1989), 352–359. 28

[SZY*09] SCHMIRGEL V., ZIMMERMANN U. E., YECHIAM E., HULIN T., PREUSCHE C.: Comprehension of operating a robot by enactive learning: Examplary approaches with programming-by-demonstration. In *Enaction on SKILLS: International Conference on Multimodal Interfaces for Skills Transfer (SKILLS '09)* (2009), pp. 135–142. 9

[TC32] TROWBRIDGE M., CASON H.: An experimental study of Thorndike's theory of learning. *Journal of General Psychology 7* (1932), 245–258. 28

[TDM*08] TÜMLER J., DOIL F., MECKE R., PAUL G., SCHENK M., PFISTER E. A., HUCKAUF A., BOCKELMANN I., ROGGENTIN A.: Mobile Augmented Reality in industrial applications: Approaches for solution of user-related issues. In *Proceedings of the 7th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '08)* (2008), IEEE Computer Society, pp. 87–90. 8

[TDTA03] TONEY A., DUNNE L., THOMAS B. H., ASHDOWN S. P.: A shoulder pad insert vibrotactile display. In *Proceedings of the 7th IEEE International Symposium on Wearable Computers (ISWC '03)* (2003), IEEE Computer Society, pp. 35–. 52

[TFCS97] TINDALL-FORD S., CHANDLER P., SWELLER J.: When two sensory modes are better than one. *Journal of Experimental Psychology: Applied 3*, 4 (1997), 257 – 287. 55

[TFR*10] TRIPICCHIO P., FILIPPESCHI A., RUFFALDI E., TECCHIA F., AVIZZANO C., BERGAMASCO M.: A measuring tool for accurate haptic modeling in industrial maintenance training. In *Haptics: Generating and Perceiving Tangible Sensations*, vol. 6192 of *Lecture Notes in Computer Science (LNCS)*. Springer Berlin / Heidelberg, 2010, pp. 377–384. 47, 48

[TGD04] TSINGOS N., GALLO E., DRETTAKIS G.: Perceptual audio rendering of complex virtual environments. *ACM Transactions on Graphics (TOG) 23* (2004), 249–258. 52

[TH09] TADAKUMA R., HOWE R.: A whole-arm tactile display system. In *Proceedings of the 3rd Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (World Haptics '09)* (2009), pp. 446–451. 59

[THDA08] TAATGEN N. A., HUSS D., DICKISON D., ANDERSON J. R.: The acquisition of robust and flexible cognitive skills. *Journal of Experimental Psychology: General 137*, 3 (2008), 548–565. 88

[TLN01] TIAN X., LING K. V., NG W. S.: Magnetic tracker calibration for an Augmented Reality system for therapy. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine 215*, 1 (2001), 51–59. 42

[Tul85] TULVING E.: How many memory systems are there? *American Psychologist 40*, 4 (1985), 385 – 398. 87

[TYN05] TANIE H., YAMANE K., NAKAMURA Y.: High marker density motion capture by retroreflective mesh suit. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '05)* (2005), pp. 2884–2889. 46

[Uni06] UNIVERSITY OF GLASGOW: Tactons, 2006. http://www.tactons.org/; website checked on May 2011. 59

[UPBS08] UNZUETA L., PEINADO M., BOULIC R., SUESCUN A.: Full-body performance animation with sequential inverse kinematics. *Graphical Models 70*, 5 (2008), 87–104. 46

[UUI99] UNDERKOFFLER J., ULLMER B., ISHII H.: Emancipated pixels: real-world graphics in the luminous room. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '99)* (1999), ACM Press/Addison-Wesley Publishing Co., pp. 385–392. 41

[vdLSB09] VAN DER LINDEN J., SCHOONDERWALDT E., BIRD J.: Good vibrations: Guiding body movements with vibrotactile feedback. In *Proceedings of the 3rd International Workshop on Physicality* (2009), pp. 13–18. 59

[VIK*02] VLAHAKIS V., IOANNIDIS N., KARIGIANNIS J., TSOTROS M., GOUNARIS M.: Virtual Reality and information technology for archaeological site promotion. In *In Proceedings of the 5th International Conference on Business Information Systems (BIS '02)* (2002), pp. 24–25. 39

[VKGM02] VIA D. K., KYLE R. R., GEIGER P. G., MONGAN P. D.: A head mounted display of anesthesia monitoring data is of value and would be used by a majority of anesthesia providers. *Anesthesia and Analgesia 95* (2002), 132. 30

[vMS05] VAN MERRIËNBOER J. J. G., SWELLER J.: Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review 17*, 2 (2005), 147–177. 55

[VTR91] VARELA, THOMPSON E., ROSCH E.: *The embodied mind: cognitive science and human experience*. MIT Press, 1991. 19

[Wag07] WAGNER D.: *Handheld Augmented Reality*. PhD thesis, Graz University of Technology, 2007. 40

[WBE*11] WEBEL S., BOCKHOLT U., ENGELKE T., GAVISH N., TECCHIA F.: Design recommendations for Augmented Reality based training of maintenance skills. In *Recent Trends of Mobile Collaborative Augmented Reality Systems*, 1st ed. Springer-Verlag, 2011, pp. 61–74. 59, 68

[WDF*10] WEBER B., DEML B., FRIEDRICH M., SCHÄTZLE S., OBERHEID H., PREUSCHE C.: Vibrotaktiles Feedback zur Aufmerksamkeitslenkung bei komplexen Lotsentätigkeiten [vibrotactile feedback for attention direction in a complex air traffic controlling task]. *Innovative Interaktionstechnologien für Mensch-Maschine-Schnittstellen* (2010), 71–85. 59

[Web08a] WEB3DCONSORTIUM: X3D, 2008. http://web3d.org/x3d/; website checked on May 2011. 3, 104

[Web08b] WEBEL S.: *Annotationen in immersiven Umgebungen: Interaktionskonzepte und Paradigmen*. VDM Verlag Dr. Müller, 2008. 68

[WH99] WICKENS C. D., HOLLANDS J. G.: *Engineering Psychology and Human Performance*, 3rd ed. Prentice Hall, 1999. 31, 32

[Wic08] WICKENS C. D.: Multiple Resources and Mental Workload. *Human Factors 50*, 3 (2008), 449–455. 55

[Wil01] WILSON M.: The case for sensorimotor coding in working memory. *Psychonomic Bulletin & Review 8*, 1 (2001), 44. 19

[WKD03] WEIDENHAUSEN J., KNOEPFLE C., D. S.: Lessons learned on the way to industrial Augmented Reality applications, a retrospective on ARVIKA. *Computers and Graphics 27*, 6 (2003), 887–891. 8

[WKZ08] WEBEL S., KEIL J., ZOELLNER M.: Multi-touch gestural interaction in X3D using hidden Markov models. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '08)* (2008), ACM, pp. 263–264. 110

[Woo99] WOODWORTH R. S.: The accuracy of voluntary movement. *The Journal of Nervous and Mental Disease 26*, 12 (1899), 743–752. 23, 31

[WSB09] WEBEL S., STAYKOVA Y., BOCKHOLT U.: Towards workflow acquisition of assembly skills using hidden Markov models. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '09)* (2009), pp. 841–846. 77, 78

[WSH*11] WEBER B., SCHÄTZLE S., HULIN T., PREUSCHE C., DEML B.: Evaluation of a vibrotactile feedback device for spatial guidance, 2011. to appear in: WorldHaptics '11. 59, 87

[WVS05] WUEST H., VIAL F., STRICKER D.: Adaptive line tracking with multiple hypotheses for Augmented Reality. In *Proceedings of the 4th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '05)* (2005), IEEE Computer Society, pp. 62–69. 43, 44

[WWK11] WIENTAPPER F., WUEST H., KUIJPER A.: Reconstruction and accurate alignment of feature maps for augmented reality, 2011. accepted for IEEE Proceedings of the Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIM-PVT). 74

[WWPR00] WEIDENBACH M., WICK C., PIEPER S., REDEL D. A.: Augmented Reality in echocardiography. a new method of computer-assisted training and image processing using virtual and real three-dimensional data sets. *Zeitschrift Für Kardiologie 89*, 3 (2000), 168–175. 10

[YEYG04] YECHIAM E., EREV I., YEHENE V., GOPHER D.: Melioration and the transition from touch typing training to everyday use. *Human Factors 45* (2004), 671–684. 29

[YGY10] YUVILER-GAVISH N., YECHIAM E.: The disadvantageous but appealing use of visual guidance in procedural skills training. *Advances in Cognitive Ergonomics* (2010), 764–773. 90

[YKY*00] YABUKAMI S., KIKUCHI H., YAMAGUCHI M., ARAI K. I., TAKAHASHI K., ITAGAKI A., WAKO N.: Motion capture system of magnetic markers using three-axial magnetic field sensor. *IEEE Transactions on Magnetics 36*, 5 (2000), 3646–3648. 47

[YMG09] YOON S. M., MALERCZYK C., GRAF H.: Skill measurement through real-time 3D reconstruction and 3D pose estimation. In *Enaction on SKILLS: International Conference on Multimodal Interfaces for Skills Transfer (SKILLS '09)* (2009), pp. 59–65. 46

[YMY01] YAMADA D., MAENO T., YAMADA Y.: Artificial finger skin having ridges and distributed tactile sensors used for grasp force control. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (2001), vol. 2, pp. 686–691. 48