

Evaluation von Situation Awareness Training in der Flugsicherung

Bruder, C., Leuchter, S., Schulze-Kissing, D. & Urbas, L.

MoDyS Research Group, Zentrum Mensch-Maschine-Systeme, TU Berlin

Einleitung

Am Zentrum Mensch-Maschine-Systeme der TU Berlin wurde ein Forschungsprototyp eines Situation Awareness Training Systems für die (*en-route*) Flugsicherung (SAT/ATC) entwickelt (Leuchter & Jürgensohn 2001). Im folgenden Bericht werden das Konzept zu dessen Evaluation und erste Ergebnisse vorgestellt: Im ersten Abschnitt wird beschrieben, auf welchen Konzepten das Trainingssystem basiert. Danach wird das Versuchsdesign und die Operationalisierung der Hypothesen der Evaluation präsentiert. Im dritten Abschnitt werden erste Ergebnisse vorgestellt. Es schließen sich die Diskussion der Ergebnisse und ein Ausblick auf Begrenzungen des Versuchsdesigns und seine Übertragbarkeit auf Situation Awareness Training in der Prozesstechnik an.

Modellbasiertes Situation Awareness Training

Situation Awareness

Situation Awareness (SA) ist handlungsrelevantes Wissen über die aktuelle Situation in einer dynamischen Aufgabenumgebung und der Prozess es zu erwerben und aktuell zu halten. Durch die Dynamik solch einer Aufgabe liegt die Handlungsrelevanz in der Gegenwart ebenso wie in der Zukunft.

Endsley & Smolensky (1998) definieren folglich SA in der Flugsicherung auf drei Ebenen:

- Ebene 1: Wahrnehmung der Elemente der Umgebung (insb. Luftfahrzeug-Daten auf dem Radarschirm),
- Ebene 2: Verständnis der Situation als Menge von Konstellationen der Elemente und
- Ebene 3: Projektion der Zustände der Elemente in die Zukunft (Trajektorien der kontrollierten Luftfahrzeuge)

SA ist somit eine komplexe Wissensstruktur, die nur unzureichend messbar ist. Die Messung des Maßes von SA erfolgt in der Praxis durch Gedächtnistests, welche die reguläre Bedienung unterbrechen (bei Flugsicherung z.B. durch *freeze* einer Verkehrssimulation und Einblenden von entsprechenden Fragen). Die Tests beziehen sich insbesondere auf Ebene 1 (z.B. Mogford 1997). Zusätzlich können auch die Ebenen 2 und 3 z.B. wie in der Methode SAGAT (Endsley 2000) durch indirekte Fragen abgedeckt werden: „*Which aircraft will need a new clearance to achieve landing requirements?*“

Mogford (1997) zeigt, dass SA auf Level 1 in der Flugsicherung (insbesondere Wissen über Höhe und Richtung der Luftfahrzeuge) ein guter Indikator für die Handlungskompetenz bei Trainingsabschluss von Fluglotsenschülern ist. Deshalb ist es nützlich SA zu trainieren.

Training

Konventionelle Trainingssysteme werden als verknüpfte mehr oder weniger dynamisch sequenzierte Abfolge von Bildschirmseiten (*screens*) umgesetzt. Um solche Systeme zu entwickeln, muss deren Abfolge vorher geplant werden. Dabei müssen alle Interventionsmöglichkeiten der Trainees berücksichtigt werden. Bei dynamischen und komplexen Mensch-Maschine-Systemen ist das nicht mehr möglich, weil die Bediener zu jedem beliebigen Zeitpunkt eine Vielzahl von Eingriffsmöglichkeiten haben, die jede wieder zu unterschiedlichen Folgesituationen führt. Deshalb werden Lehrsysteme, in denen die Bedienung von Mensch-Maschine-Systemen trainiert wird, zumeist als Simulationen umgesetzt (Chappell & Mitchell 1997; Loftin & Kenney 1995; Götte, Urbas & Wozny 1998).

In Simulationstrainingssystemen soll Handlungswissen vermittelt werden. SA hingegen ist als Zustand durch die Qualität des mentalen Situationsabbildes („*picture*“) definiert. SA-Training soll die Ausbildung von Strategien zur effizienten Auswahl von Objektinformation sowie deren Integration unterstützen, damit das *picture* aufrechterhalten werden kann.

Aufmerksamkeitssteuerung

Benutzt wird ein Simulationstraining, in dem als zusätzliche Trainingsunterstützung ein Attention-Guidance-Ansatz mit wechselnden Farbmarkierungen (Bass 1998) umgesetzt wurde. Für die Steuerung der Farbcodierung wird eine simulierte gewichtete „mentale“ Repräsentation der Verkehrssituation genutzt. Sie resultiert aus einer parallel-laufenden Simulation eines kognitiven Modells erfahrener Fluglotsen (Niessen & Eyferth 2001). Der aktuelle Zustand des simulierten *pictures* wird durch Farbcodierungen so im Radarbild dargestellt, dass die Aufmerksamkeitsverteilung zwischen den räumlich dargestellten Elementen der Situation gesteuert wird. Ergebnisse von Gopher (1993) zeigen, dass Aufmerksamkeitssteuerung im Training einen positiven Einfluss auf die Handlungskompetenz von Pilotenschülern hat.

Handlungskompetenz

Das evaluierte Trainingssystem hat die Flugsicherung im *en-route*-Bereich zum Inhalt. Hier überwachen jeweils zwei Lotsen einen Luftraumsektor. Es gilt den Verkehr sicher und wirtschaftlich zu koordinieren. Insbesondere müssen gefährliche Annäherungen zwischen Flugzeugen verhindert werden. Der Verkehr ist zwar im Voraus geplant, doch durch äußere Einflüsse wie das Wetter kann die langfristige Planung nicht vollständig umgesetzt werden. Der *Planungslotse* ist für die mittelfristige Koordination mit den Nachbarsektoren zuständig. Dazu benutzt er hauptsächlich „Flugstreifen“, auf denen geplante Routen und Überflugzeitpunkte vermerkt sind. Das wichtigste Instrument des *Exekutivlotsen* ist hingegen der Radarschirm, auf dem der Verkehr im überwachten Sektor dargestellt wird. Er ist für einen konfliktfreien Verkehr im Sektor und die Übergabe in den Folgesektor verantwortlich. Dazu gibt er per Sprechfunk Anweisungen zu Kurs-, Geschwindigkeits- oder Höhenänderungen an die Piloten. Das Trainingssystem unterstützt die Simulatoreausbildung des Exekutivlotsen.

Die Verbesserung der Strategien zur Bildung und Aufrechterhaltung des *pictures* wird als Erfolg bei der Aufmerksamkeitssteuerung interpretiert und an sekundären Indikatoren wie Gedächtnistest, Fehleranzahl und Geschwindigkeit bei der Behandlung von Ein- und Ausfliegern in den Sektor (Sekundäraufgabe), der Konflikterkennung und -lösung (Erkennen möglicher Folgekonflikte) gemessen.

Evaluation von SAT/ATC

Die Evaluation des vorgestellten modellgestützten Simulationstraining SAT/ATC zielt auf die Beantwortung der Frage ab, welche Effekte SAT/ATC auf die Aufmerksamkeitssteuerung, Handlungskompetenz und Beanspruchung hat. Zum Vergleich wurde das Training ohne SAT/ATC durchgeführt.

Im Rahmen eines Simulationsexperimentes (Bildschirmgestaltung s. Abbildung 1) wurden die wesentlichen Systemeigenschaften des Mensch-Maschine-Systems zur experimentellen Untersuchung im Labor nachgebildet.

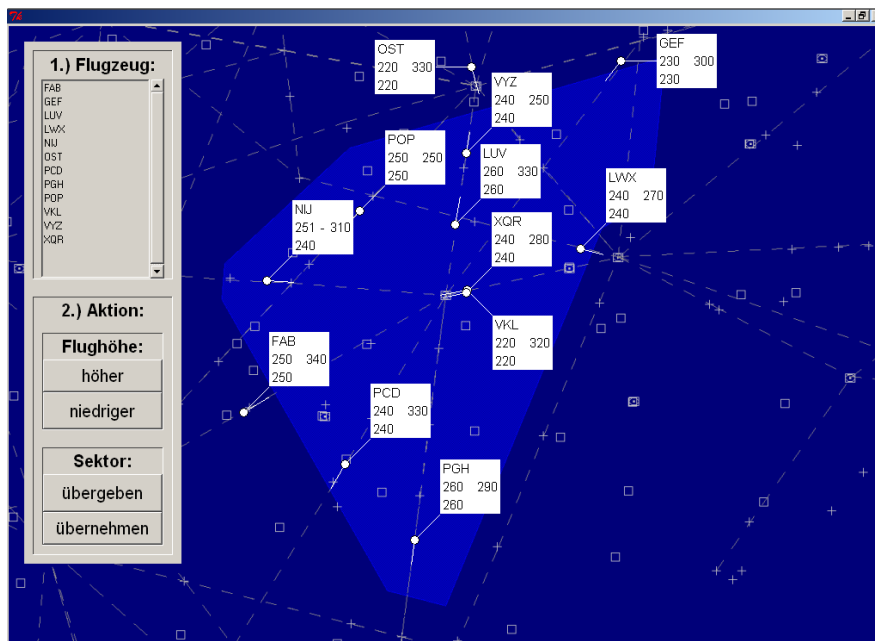


Abbildung 1: An die Evaluation angepasstes modellbasiertes Situation Awareness Training SAT/ATC

Einführung in das Simulationsexperiment

Das Simulationsexperiment beinhaltet aus zeitlichen und materiellen Gründen nicht die Umsetzung des Gesamtsystems, stattdessen werden hypothesengeleitet für die Beantwortung der Fragestellung wesentliche Systemeigenschaften nachgebildet. Im Zentrum steht nicht die Prüfung des Funktionierens technischer Eigenschaften, sondern die Bewertung der Gestaltungsgüte auf das Ausführen der Arbeitsaufgaben der zukünftigen Nutzer bzw. Trainierenden.

Das Besondere bei der Simulation im Vergleich zu anderen Bewertungsmethoden ist, dass die Dynamik des Handlungsablaufs untersucht werden kann. Der Nutzer bzw. Trainierende kann prototypische Arbeits- bzw. Trainingsaufgaben im Simulationsexperiment ausführen. Nur so lässt sich entscheiden, wie stark die Arbeit am System den Nutzer belastet, welche Fehler er macht, welche Informationen er benötigt und wie lange er braucht, um seine Aufgabe zu erfüllen.

Wesentlichen Einfluss auf die Güte der Ergebnisse hat die Umsetzung der Arbeits- bzw. Trainingsaufgaben in Simulationsaufgaben. Bei der Konstruktion der Simulationsaufgaben ist es wichtig, dass die Wirkung der Trainingsaufgabe Rückschlüsse auf die kognitiven Prozesse des Trainierenden ermöglicht und repräsentativ für die realen Arbeits- bzw. Trainingsaufgaben ist.

Aufbau der Untersuchung

Kritische Bedingungsvariation

Ziel der Untersuchung ist die Bewertung von SAT/ATC.

Es werden zum einen Aussagen zum direkten Einfluss von SAT/ATC auf die Aufmerksamkeit, Handlungskompetenz und Beanspruchung angestrebt.

Dazu führt eine Gruppe von Probanden, die Versuchsgruppe, die Simulationen mit SAT/ATC durch, und eine andere Gruppe, die Kontrollgruppe, führt diese ohne SAT/ATC durch. Die Versuchsgruppe wird bei der Bearbeitung der Simulationen durch den modellierten Fluglotsentrainer unterstützt. Dieser lenkt die Aufmerksamkeit durch farbliche Markierung auf bestimmte Flugzeuge. Die Kontrollgruppe durchläuft den Versuch in gleicher Weise, wobei keine Unterstützung durch SAT/ATC erfolgt.

Zum anderen werden aber auch Aussagen zum Transfereffekt des Trainings mit SAT/ATC angestrebt, also ob der Lerneffekt übertragen wird auf Situationen, in denen SAT/ATC nicht zur Verfügung steht.

Daher trainiert die Versuchsgruppe abwechselnd mit und ohne Unterstützung durch SAT/ATC. In den Simulationen ohne SAT/ATC soll sich der Transfereffekt zeigen.

Ablauf des Versuchs

Grundlegend gliedert sich das Versuchsdesign in zwei Versuchstage, an denen jeder der 44 Probanden teilnimmt. Der erste Versuchstag (s. Tabelle 1) beinhaltet ein computerbasiertes Tutorium, in dem die Versuchsteilnehmer die wesentlichen Aufgaben eines Fluglotsen kennen lernen und Wissensfragen beantworteten. Danach üben die Probanden in einer Radarsimulation die gelernten Aufgaben.

Die Aufgabenbewältigung in der Übung wird als Kriterium für die Zuweisung der Probanden zur Kontroll- und Versuchsgruppe herangezogen. Anhand ihres Leistungsprofils in der Übung werden die Personen gleichmäßig auf die Versuchs- und Kontrollgruppe verteilt. Dadurch wird sichergestellt, dass die beiden Gruppen hinsichtlich der Ausgangsleistung gleich sind.

Der zweite Versuchstag (s. Tabelle 1) stellt das eigentliche Simulationsexperiment dar. In zwölf aufeinander folgenden Sequenzen aus Simulation, teilweise Gedächtnistest und Fragebogen zur Aufgabeneinschätzung trainiert der Proband. Eingerahmt wird das Simulationstraining durch einen Beanspruchungsfragebogen, den die Teilnehmer jeweils vor und nach den zwölf Trainingssimulationen ausfüllen. Ergänzend kommen Fragebögen und Interview zur subjektiven Einschätzung des Trainingssystems SAT/ATC zur Anwendung.

Tabelle 1: Versuchsdesign zur Evaluation von SAT/ATC im Überblick (vereinfacht)

Versuchsdesign: Grobplanung	
Phase	Beschreibung
Tag 1: Tutorium, Instruktion und Übung	
Tutorium:	Lernprogramm mit Wissensvermittlung und Wissensprüfung
Übungssimulation:	Radarsimulation von ca. 10 Minuten zum Üben der Navigation
Tag 2: Simulationsexperiment, Fragebögen und Interview	
Instruktion:	Wiederholung und Erklärung der Aufgaben für Versuchs- und Kontrollgruppe
Ausgangsbeanspruchung:	BLV (Künstler 1980) vor dem Simulationstraining
Simulationssequenzen (zwölf Wiederholungen):	Simulationssequenz aus Radarsimulation, NASA TLX (Hart & Staveland 1988) und Gedächtnistest
Trainingsbeanspruchung:	BLV nach dem Simulationstraining
Fragebögen:	Kontrollwahrnehmung im Simulationstraining und Akzeptanz
Interview:	Erfassung von Vorgehen, Strategien und Einschätzung

Konstruktion der Trainingssimulationen

Die Trainingssimulationen stellen kurze Luftraumsimulationen dar. Die Probanden sollen sich bei jeder Simulation vorstellen, dass sie unerwartet die Kontrolle über den Sektor übernehmen müssen, weil z.B. der zuständige Fluglotse ausfällt.

Die Trainingssimulationen sollen folgenden Kriterien genügen:

- Damit die Simulationen vergleichbar sind, haben alle Trainingsaufgaben die gleiche Schwierigkeit und Struktur.
- Die Simulationen sollen die kognitiven Prozesse induzieren, die SAT/ATC trainieren soll (siehe Anforderungen von SAT/ATC), d.h. den Aufbau und die Aufrechterhaltung von SA mit Wahrnehmung, Interpretation und Antizipation der Verkehrssituation.
- Das Verhalten der Probanden soll automatisch protokolliert und hypothesengeleitet auswertbar sein. Es soll ein wesentlicher Indikator für die Bewertung von SAT/ATC sein.

Eigenschaften der Simulationen

Per Instruktion wird der Proband aufgefordert, den Luftraum zu überwachen, konflikthafte Situationen zu erkennen, Kollisionen zu vermeiden, einfliegende Flugzeuge zu übernehmen und ausfliegende zu übergeben.

Die zwölf Trainingsszenarien haben die folgenden Eigenschaften:

- Es wird ein realer en-route-Sektor mit Luftfahrtstraßen verwendet.
- Pro Simulation sind zehn bis zwölf Flugobjekte zu überwachen, wobei sich immer mindestens acht im Sektor befinden.

- Die Dauer einer Simulation beträgt 4 Minuten.
- Pro Simulation ereignet sich ein Hauptkonflikt, der durch eine Veränderung der Höhe eines der im Konflikt befindlichen Flugobjekte verhindert werden kann.

Jede Konfliktsituation ist so konstruiert, dass für jedes daran beteiligte Flugzeug eine Lösung ohne Folgekonflikt und eine Lösung mit Folgekonflikt möglich ist.

Ein Konflikt besteht darin, dass auf gleicher Höhe (Konfliktarten):

- zwei Flugzeuge einander entgegen fliegen,
- zwei Flugzeuge sich kreuzen oder
- ein schnelleres hinter einem langsameren fliegt.

Der Ort des Konfliktes und die Konfliktart wurden systematisch kombiniert.

Es sind sowohl Level- als auch Steig-Sinkflug-Konflikte enthalten.

Für jedes Szenario wurden auch Scheinkonflikte konstruiert. Scheinkonflikt meint, dass zwar ein oder zwei Flugzeuge sich bzw. die Flugzeuge des Hauptkonfliktes kreuzen, entgegenfliegen oder überholen, aber nicht auf gleicher Höhe wie die Konfliktpartner fliegen. Durch Vergleichsprozesse können sie als Konflikte ausgeschlossen werden.

Anhand dieser Kriterien werden zwölf Szenarien konstruiert, von denen acht jeweils einen Konflikt, zwei jeweils zwei Konflikte und zwei Distraktoren gar keinen Konflikt enthalten.

Hypothesen und Operationalisierung

Da die Anforderungen an SAT/ATC aus dem Situation-Awareness-Konzept von Endsley & Smolensky (1998) abgeleitet sind, wird die Evaluation von SAT/ATC untersuchen, ob die Kriterien von SA tatsächlich durch den Einsatz von SAT/ATC trainiert werden.

Kriterien für SA lassen sich nach Endsley & Smolensky (1998) auf drei Ebenen finden:

- Die Ebene 1 beinhaltet das Wissen über den aktuellen Zustand der Objekte, d.h. im Bereich der Flugsicherung die Wahrnehmung der Flugzeuge und ihrer Informationen.
- Die Ebene 2 enthält das Wissen über die Bedeutung der Objekte für die Situation, d.h. die Interpretation der wahrgenommenen Flugzeuge und ihrer Informationen.
- Die Ebene 3 problematisiert das Wissen über den zukünftigen Zustand der Objekte, d.h. die Prognose der Position bzw. des Status der Flugzeuge.

Hypothesen zur Bewertung von SAT/ATC

Bei der Beantwortung der Fragestellung zu den Effekten des Trainingssystems SAT/ATC auf die Aufmerksamkeit, Handlungskompetenz und Beanspruchung wird folgende allgemeine Hypothese formuliert:

SAT/ATC fördert den Aufbau von SA, d.h. die Probanden der Versuchsgruppe sind im Vorteil gegenüber den Probanden der Kontrollgruppe im Hinblick auf:

- die Aufmerksamkeitssteuerung
- die Konflikterkennung und
- die Konfliktlösung.

Diese generelle Hypothese wird in spezielle Hypothesen (s. Tabelle 2, auszugsweise für Aufmerksamkeitslenkung und Handlungskompetenz) überführt.

Tabelle 2: Spezielle Hypothesen zur Aufmerksamkeit und Handlungskompetenz

SAT/ATC lenkt die Aufmerksamkeit auf kritische Luftraumsituationen:	
→	Die Probanden der Versuchsgruppe beachten häufiger kritische Flugzeuge als die Probanden der Kontrollgruppe (Aufmerksamkeitsparameter)
→	Die Versuchsgruppe erinnert sich häufiger an kritische Flugzeuge als die Kontrollgruppe (Gedächtnisparameter).
SAT/ATC unterstützt die Erkennung von Konfliktsituationen:	
→	Die Versuchsgruppe nimmt die Konfliktpartner zusammenhängend wahr (Aufmerksamkeitsparameter).
→	Die Versuchsgruppe erkennt häufiger die Konfliktpartner als die Kontrollgruppe (Kompetenzparameter).
→	Die Versuchsgruppe erkennt die Konfliktpartner schneller als die Kontrollgruppe (Kompetenzparameter).
SAT/ATC unterstützt die Lösung von Konflikten:	
→	Die Versuchsgruppe löst die Konflikte häufiger als die Kontrollgruppe (Kompetenzparameter).
→	Die Versuchsgruppe löst die Konflikte schneller als die Kontrollgruppe (Kompetenzparameter).
→	Die Versuchsgruppe lösen die Konflikte häufiger ohne Folgekonflikte (Kompetenzparameter).

Operationalisierung

Um die Kriterien für SA zu messen, kommen die folgende Indikatoren zum Einsatz (s. Tabelle 3):

- Das Wissen über den aktuellen Zustand (operationalisiert durch die Erinnerungsleistung am Ende einer Simulation) wird durch Gedächtnistest erhoben (Indikator für Kriterium auf Ebene 1).
- Das Wissen um die Bedeutung der Objekte (operationalisiert über die Zuweisung von Aufmerksamkeit auf bedeutsame Objekte) wird durch Blickanalysen gemessen (Indikator für Kriterium auf Ebene 2).
- Die Erkennung von Konflikten (operationalisiert als Zeitpunkt der Markierung eines Objektes durch die Versuchsperson) wird durch Latenzzeitmessungen erhoben (Indikator für Kriterium auf Ebene 2)
- Die Fähigkeit zur Prognose zukünftiger Flugpositionen (operationalisiert über Anzahl erzeugter Folgekonflikte) wird anhand von Fehlerhäufigkeiten gemessen (Indikator für Kriterium auf Ebene 3)
- Aufgabenbelastung und Beanspruchung werden über Beurteilungsverfahren zur psychischen Beanspruchung (*cognitive workload*) erhoben.
- Fragebögen und Interview dienen der Beurteilung von Akzeptanz und Kontrollerleben.

Tabelle 3: Ableitung der verwendeten Indikatoren aus Kriterien von SA (Endsley & Smolensky 1998)

Kriterium für SA		Indikator von SAT/ATC
Wahrnehmung der Luftraumsituation (Ebene 1)	→	Blickanalysen, Gedächtnistest
Erkennung von Konflikten (Ebene 2)	→	Häufigkeit und Zeit bis zur Erkennung
Prognose zukünftiger Flugpositionen und Behebung von Konflikten (Ebene 3)	→	Häufigkeit bis Konfliktlösung, Zeit bis zur Konfliktbehebung, Güte der Lösung

Die Umsetzung der Indikatoren in messbare Parameter von SA in SAT/ATC erfolgt über die Erhebung der hier aufgeführten abhängigen Variablen (s. Tabelle 4):

- Blickdaten der Simulationen, wie z.B. Fixationshäufigkeiten,
- Handlungsdaten der Simulationen, wie z.B. Fehler und Latenzzeiten,
- Gedächtnisdaten, z.B. erinnerte Flughöhe und
- Belastungsdaten, wie z.B. Aufgabenbeanspruchung.

Tabelle 4: Zuordnung der verwendeten Parameter zu den Indikatoren der Evaluation

Indikatoren der Evaluation		Parameter
Blickanalyse	→	Fixationshäufigkeit, Übergangswahrscheinlichkeit
Gedächtnistest	→	Cued-Recall von Höhe, Speed und Richtung
Erkennung der Konflikte	→	Fehlerhäufigkeit und Latenzzeit
Behebung der Konflikte	→	Fehlerhäufigkeit, Latenzzeit und Güte
Fragebögen	→	Aufgabenbelastung und Akzeptanz

Messmethoden

Wie aus dem vorigen Abschnitt zur Operationalisierung der Hypothesen deutlich wird, kommen verschiedene Messmethoden im Rahmen der Evaluation zur Anwendung. Im folgenden Abschnitt werden die Methode der Aufmerksamkeitsmessung, der Handlungskompetenzmessung sowie der Erinnerungsmessung kurz vorgestellt.

Aufmerksamkeitsmessung

Die visuelle Aufmerksamkeit als Indikator für die Wahrnehmung und Interpretation der Verkehrssituation wird über Blickanalysen erhoben. Entscheidender Parameter ist die Fixationshäufigkeit einzelner Flugzeuge. Die Probanden tragen während der Bearbeitung der Simulationen einen kopfbasierten *eye-tracker*. Diese Blickbewegungsmessanlage der Firma SMI ist gekennzeichnet durch eine zeitliche Auflösung von 50 Hz und durch eine örtliche Auflösung von 0,5° bis 1,0°. Die Einbeziehung der Kopfbewegungen in die Berechnung der Blickdaten wird durch einen elektromagnetischen Kopfbewegungstracker ermöglicht. Die Blickkoordinaten werden während der Simulation aufgezeichnet und können anschließend automatisiert ausgewertet werden. Durch die Synchronisation von Simulationsdaten und Blickdaten erfolgt die Zuordnung von Fixationen der Probanden zu Objekten auf dem Bildschirm.

Handlungskompetenzmessung

Die Handlungskompetenz als Parameter für die Prognose der Verkehrssituation und dadurch die Lösung konfliktbehafteter Situationen wird mit Hilfe von Logfile-Protokollierung der Aktionen des Benutzers als auch der Aktionen von SAT/ATC gemessen. Die Handlungskompetenz betrifft die Erkennung und Behebung des eigentlichen Konfliktes, die Übernahme und Übergabe der Ein- und Ausflieger, überzählige Eingriffe und die Erkennung und Behebung des Zweitkonfliktes.

Messung der Erinnerungsleistung

Die Erhebung der Erinnerungsleistung bezüglich luftfahrzeugbezogener Parameter liefert Daten, welche als Indikator für das Wissen um den aktuellen Zustand der Objekte (Ebene 1) im Moment des Simulationendes dienen.

Nach dem Ende der Szenarienbearbeitung erhält der Proband die Aufgabe, anhand einer Darstellung den Positionen der zuletzt kontrollierten Luftfahrzeuge im Sektor die entsprechenden Parameter zuzuordnen. Zur Eingabe der erinnerten Parameter für ein Luftfahrzeug muss der Proband jeweils die markierte Position per Mouse-Click auswählen. Nun besteht über ein links auf dem Bildschirm erscheinendes Fenster die Möglichkeit zur Eingabe der Flughöhe und Fluggeschwindigkeit in jeweils hierfür vorgesehene Masken. Über die Darstellung einer Windrose kann darüber hinaus die Angabe der Flugrichtung vorgenommen werden.

Durchführung der Untersuchung

Die Untersuchung wurde im Juni und Juli 2002 mit 44 Probanden im Labor des Zentrums Mensch-Maschine-Systeme (ZMMS) der Technischen Universität Berlin durchgeführt. Die Probanden sind Abiturienten, Studenten und Auszubildende. Sie entsprechen hinsichtlich Alter und Bildung den Anwärtern für eine Fluglotsenausbildung.

Der Versuch wurde für jede Person einzeln und an zwei Versuchstagen, die etwa fünf bis acht Tage auseinander liegen, durchgeführt.

Vorläufige Ergebnisse

Es werden die Ergebnisse zur Handlungskompetenz, Beanspruchung und Akzeptanz berichtet. Die Ergebnisse zur Gedächtnisleistung und zur Aufmerksamkeit werden gerade ausgewertet und liegen noch nicht vor.

Ergebnisse zu den Parametern der Handlungskompetenz

Die Handlungskompetenz wird hinsichtlich folgender Aspekte ausgewertet:

- Trendschätzung über die zehn konfliktbehafteten Simulationen zur Abschätzung des Trainingseffektes durch die Trainingssimulation
- Vergleich von Versuchs- und Kontrollgruppe zur Beurteilung des Effektes von SAT/ATC im Vergleich zum Training ohne SAT/ATC
- Vergleich der Simulationen der Versuchsgruppe mit und ohne SAT/ATC zur Prüfung des Transfereffektes durch SAT/ATC.

Es wurden hinsichtlich der Variablen Erkennung des Hauptkonfliktes, Lösung des Hauptkonfliktes, Übernehmen der Einflieger und Übergeben der Ausflieger, überzählige Eingriffe und Erkennung und Lösung des Zweitkonfliktes hauptsächlich Latenzzeiten und Fehlerhäufigkeiten ausgewertet.

Die Versuchsgruppe und Kontrollgruppe setzen sich jeweils aus 22 Versuchsteilnehmern zusammen. Im Folgenden werden die wichtigsten Ergebnisse berichtet.

Erkennung der Konflikte

Durch Markierung der beteiligten Konfliktpartner wird die Erkennung des Konfliktes erhoben. Sowohl die Anzahl der **nicht** markierten Konfliktpaare (Fehlerhäufigkeit) als auch die Latenzzeit der Markierung (Zeit vom Start der Simulation bis zur Markierung) wurde erhoben, ausgewertet und statistisch bewertet.

Die Häufigkeit **nicht** markierter Konfliktpaare ist mit zehn für die Versuchsgruppe und fünf für die Kontrollgruppe zwar zu gering für eine statistische Beurteilung, zeigt aber eine deutlich höhere Fehlerzahl bei den Probanden der Versuchsgruppe.

Die Latenzzeit der Erkennung zeigt über alle Versuchspersonen einen linearen Trend im Sinne einer Abnahme der Erkennenszeit. Das spricht für einen Trainingseffekt durch das Simulationsexperiment. Dieser Trainingseffekt unterscheidet sich aber weder zwischen der Versuchs- und Kontrollgruppe, noch innerhalb der Versuchsgruppe zwischen den Simulationen mit und ohne SAT/ATC.

Lösung der Konflikte

Durch eine Änderung der Höhe eines der am Konflikt beteiligten Flugzeuge über das Menü des Simulationssystems kann der Konflikt gelöst werden.

Die Konflikte können nicht gelöst oder gelöst sein. Weiterhin können die Konflikte richtig gelöst sein (ohne Folgekonflikte) oder schlecht gelöst sein (mit Folgekonflikten).

Es wurden die Häufigkeiten **nicht** und **nicht richtig** gelöster Konflikte als Fehlerhäufigkeit ausgewertet (s. Abbildung 2).

Die Latenzzeit ist die Zeit vom Start der Simulation bis zum Eingriff mittels Menü des Simulationssystems.

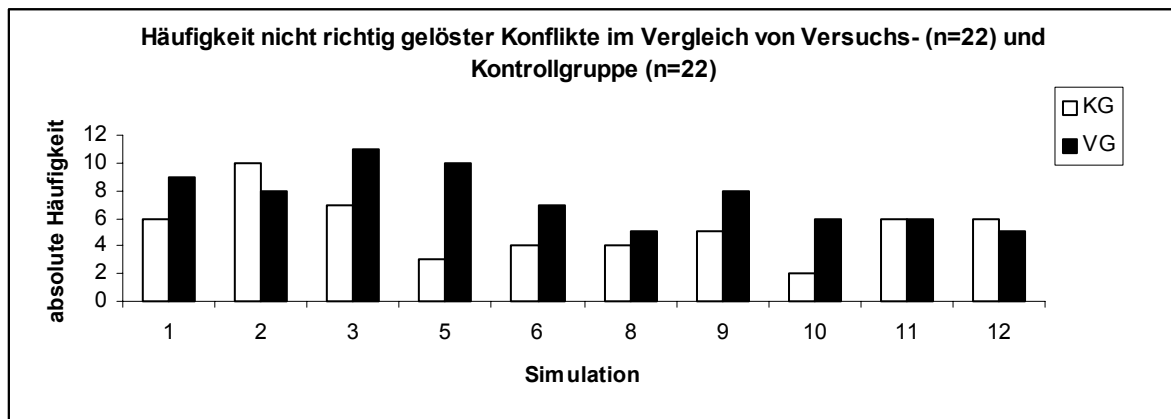


Abbildung 2: Fehlerhäufigkeiten bei der Lösung des Hauptkonfliktes

Die Fehlerhäufigkeiten zeigen über alle Versuchspersonen einen Trend in Richtung einer Abnahme der Fehlerhäufigkeit.

Vergleicht man den Unterschied der Fehlerhäufigkeiten von Versuchs- und Kontrollgruppe, so ergibt sich ein statistisch bedeutsamer Unterschied in den Simulationen fünf, sechs, neun und zehn. Das heißt, die Probanden in der Versuchsgruppe machen signifikant mehr Fehler als die der Kontrollgruppe.

Die Latenzzeiten der Konfliktlösung zeigen zwar einen linearen Trend im Sinne eines Trainingseffektes, d.h. sie nehmen systematisch im Verlauf des Experimentes ab. Es zeigt sich jedoch kein Unterschied zwischen den beiden Gruppen.

Übernahmen und Übergaben

Das Übernehmen der Einflieger und Übergeben der Ausflieger, d.h. die Simulation der Interaktion mit den Nachbarlotsen dient im Sinne einer Doppelaufgabe als Indikator für die Beanspruchung durch die Konflikterkennung und Konfliktlösung.

Die Übernahmen und Übergaben sind hinsichtlich der relativen Häufigkeit getätigter Übernahmen und Übergaben an den erwarteten Übernahmen und Übergaben ausgewertet.

Es wurden 50 bis 100% der erwarteten Übernahmen und Übergaben in den jeweiligen Simulationen durchgeführt.

Es zeigen sich weder ein Trend über die Simulationen, noch ein Unterschied zwischen den Gruppen hinsichtlich der Übernahme- und Übergabehäufigkeit.

Überzählige Eingriffe

Überzählige Eingriffe sind alle Eingriffe, bei denen die gleiche Aktion wiederholt ausgeführt wird als auch alle unnötigen Aktionen.

Es sind die mittlere Häufigkeiten der überzähligen Eingriffe ausgewertet.

Es zeigt sich ein linearer Trend über alle Versuchsteilnehmer, der die systematische Abnahme der überzähligen Eingriffe widerspiegelt.

Weiterhin ist ein signifikanter Unterschied zwischen Versuchs- und Kontrollgruppe nachweisbar. Dieser wird darin sichtbar, dass im Mittel in der Versuchsgruppe mehr zusätzliche Eingriffe vorgenommen wurden als in der Kontrollgruppe (s. Abbildung 3).

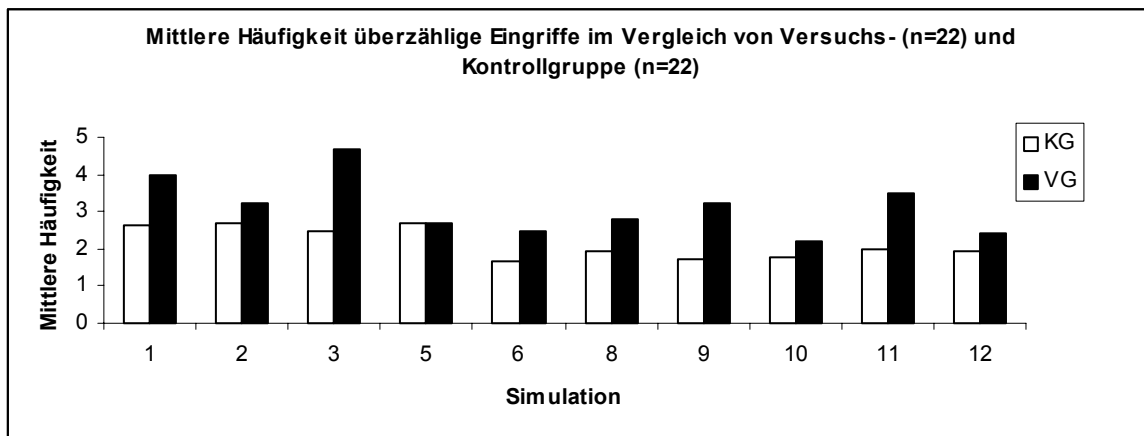


Abbildung 3: Mittlere Häufigkeit überzähliger Eingriffe

Diese überzähligen Eingriffe sind der Gruppe der doppelt getätigten Eingriffe zuzuordnen.

Vergleicht man die überzähligen Eingriffe der Versuchsgruppe paarweise in Simulationen mit und ohne SAT/ATC, so zeigen sich robuste Unterschiede zwischen den Simulationen drei (mit SAT/ATC) und sechs (ohne SAT/ATC) und zwischen den Simulationen elf (mit SAT/ATC) und zwölf (ohne SAT/ATC). Die Probanden der Versuchsgruppe haben in den Simulationen mit Unterstützung durch SAT/ATC mehr überzählige Eingriffe getätigt als in den Simulationen ohne SAT/ATC (s. Abbildung 4).

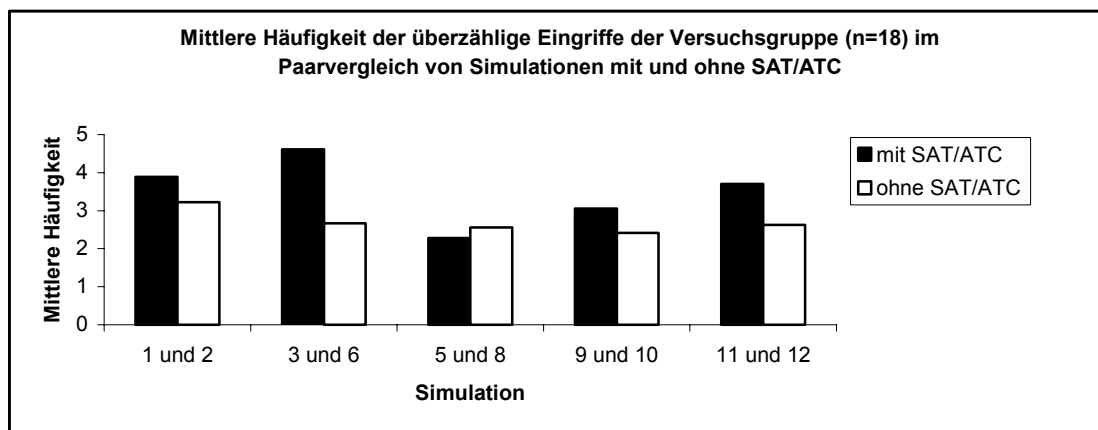


Abbildung 4: Paarvergleich der mittleren Häufigkeiten der überzähligen Eingriffe der Versuchsgruppe

Lösung und Erkennung des Zweitkonfliktes

In zwei der zwölf Szenarien haben die Versuchspersonen im Unterschied zu den anderen zehn Szenarien mehr als ein Konflikt zu bewältigen. Der Zweitkonflikt ist in Art und Lokalisation vergleichbar mit dem Hauptkonflikt.

Die Erkennung und Lösung des Zweitkonfliktes sind hinsichtlich Fehlerhäufigkeit und Latenzzeit ausgewertet.

Die Fehlerhäufigkeit der Erkennung des Zweitkonfliktes ist in beiden Szenarien mit Zweitkonflikt in der Versuchsgruppe signifikant höher als in der Kontrollgruppe (s. Abbildung 5).

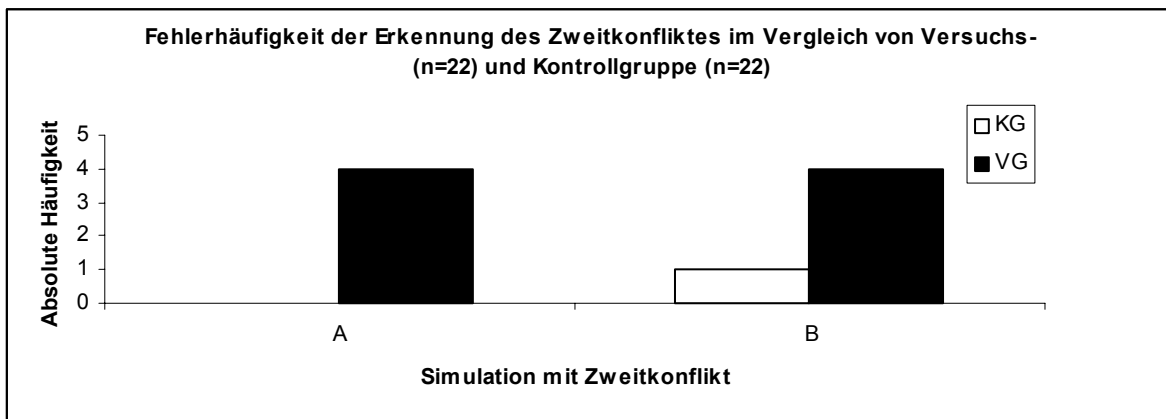


Abbildung 5: Fehlerhäufigkeit der Erkennung des Zweitkonfliktes

Die Häufigkeit der **nicht** und **nicht richtig** gelösten Zweitkonflikte ist ebenfalls in der Versuchsgruppe signifikant höher als in der Kontrollgruppe (s. Abbildung 6).

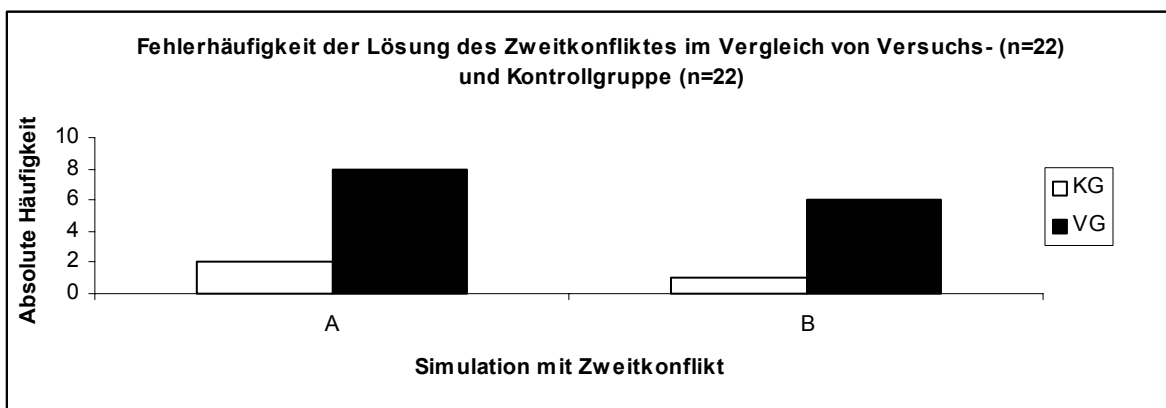


Abbildung 6: Fehlerhäufigkeit der Lösung des Zweitkonfliktes

Ergebnisse zur Beanspruchung und Akzeptanz

Um die Beanspruchung durch die Simulationsaufgaben zu erfassen und den Effekt von SAT/ATC auf diese zu untersuchen, wenden wir folgende Verfahren an:

- Der NASA TLX (Hart & Staveland 1988) ermöglicht die Erfassung der durch die Simulationen bedingten Beanspruchung und wird nach jeder Simulation erhoben.
- Der BLV (Künstler 1980) wird eingesetzt zur Erhebung der Beanspruchung vor und nach dem Simulationsexperiment.

Zur Einschätzung der Akzeptanz von SAT/ATC werden Fragebogen und Interview am Ende des Versuchs durchgeführt.

Ergebnisse zur Beanspruchung

Die Auswertung des BLV ergibt eine bedeutsame Zunahme der Beanspruchung im Verlauf des Simulationsexperimentes, d.h. die Anstrengung und Ermüdung sind gestiegen. Jedoch existiert kein bedeutsamer Unterschied zwischen der Versuchs- (n=22) und Kontrollgruppe (n=22).

Die Auswertung des NASA TLX ergibt eine Abnahme der durch die Simulation bedingten Aufgabenbeanspruchung. Augenscheinlich sind die Werte der Versuchsgruppe in nahezu jeder Simulation höher als die Werte der Kontrollgruppe, dieser Unterschied hinsichtlich einer höheren Aufgabenbeanspruchung der Versuchsgruppe ist aber nicht statistisch bedeutsam (s. Abbildung 7).

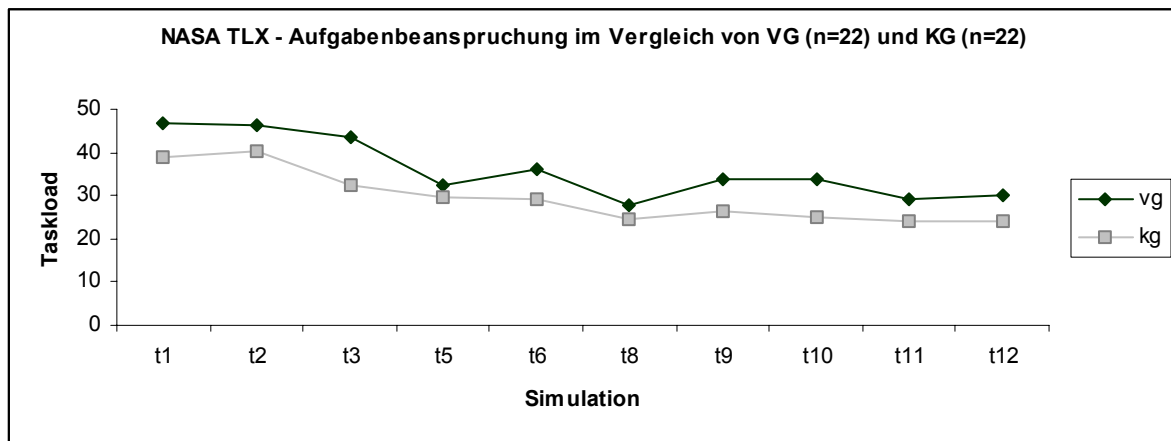


Abbildung 7: Aufgabenbeanspruchung erhoben mit dem NASA TLX

Beurteilung von SAT/ATC durch die Versuchsteilnehmer

Am Ende der Darstellung der vorläufigen Ergebnisse werden ergänzend die wichtigsten Befunde der Fragebögen und des Interviews benannt.

Die Ergebnisse beziehen sich nur auf die Versuchsgruppe, da nur sie Erfahrung mit SAT/ATC gesammelt hat.

SAT/ATC wird im Hinblick auf Farbwahl, Bedienung und emotionaler Bewertung als gut eingeschätzt.

Das Interview zu SAT/ATC ergibt, dass die Versuchspersonen die Farben als weitestgehend gut gewählt einschätzen, den Hinweisen von SAT/ATC trauen und sich durch SAT/ATC nicht bevormundet gefühlt haben. Die Versuchsteilnehmer geben aber auch einige Probleme mit SAT/ATC an (auszugsweise):

- Die Hinweise kommen oft erst, nachdem der Versuchsteilnehmer den Konflikt identifiziert und markiert hat.
- Die Farben ändern sich verzögert und unregelmäßig, nachdem der Versuchsteilnehmer einen Eingriff getätigt hat. Das führt zur Verunsicherung, so dass die betreffende Person die Aktion überflüssigerweise wiederholt.

Integrative Diskussion der Ergebnisse

Zusammenfassend zeigen die vorläufigen Ergebnisse einen robusten Lerneffekt durch das Simulationstraining. Dieser Lerneffekt zeigt sich nahezu durchgängig sowohl in der Abnahme der Fehlerhäufigkeiten als auch in der Verringerung der Latenzzeiten.

Im Hinblick auf den Aufbau von SA durch SAT/ATC deuten die berichteten Ergebnisse eher an, dass das evaluierte SA Training der Flugsicherung keinen, teilweise einen hinderlichen Einfluss auf den Aufbau von SA hat. Die vorläufigen Ergebnisse weisen entweder keine Unterschiede zwischen Versuchs- und Kontrollgruppe oder Nachteile der mit SAT/ATC trainierenden Versuchsgruppe auf.

Insbesondere hinsichtlich der Latenzzeiten ist kein Effekt von SAT/ATC aufzufinden.

Wie in den berichteten Ergebnissen ersichtlich wird, existieren in einigen Variablen und ihren Parametern bedeutsame Unterschiede zwischen Versuchs- und Kontrollgruppe.

Diese Unterschiede deuten eine Behinderung des Aufbaus von SA der Ebene 3 an. Diese Schlussfolgerung wird gestützt durch die höhere Fehlerhäufigkeit der Lösung des Hauptkonfliktes, die höhere Häufigkeit überzähliger Eingriffe und die erhöhte Fehlerhäufigkeit bei der Erkennung und Lösung des Zweitkonfliktes. Das Training mit SAT/ATC scheint teilweise hinderlich auf die Prognosefähigkeit und somit auch auf die Handlungskompetenz der Versuchsteilnehmer zu wirken.

Diese verminderte Prognosefähigkeit der Versuchsgruppe gegenüber der Kontrollgruppe ist nicht durch eine geringere Beanspruchung der Versuchsgruppe erklärbar. Vielmehr nimmt die allgemeine Beanspruchung (BLV) während des Simulationstrainings bedeutsam zu. Die Befunde zur Aufgabenbeanspruchung (NASA TLX) zeigen außer einer allgemeinen Abnahme eine durchgängig leicht höhere Beanspruchung beim Training mit SAT/ATC.

Besonders hervorzuheben sind die Ergebnisse, dass SAT/ATC zur erhöhten Fehlerhäufigkeit beim

Zweitkonflikt führt. Diese Tatsache kann man im Zusammenhang mit den Interviewdaten der Versuchsgruppe sehen, die angeben, dass SAT/ATC zwar zur Arbeitserleichterung, aber auch zu einer automatischen Aufmerksamkeitslenkung führt. Statt einer besseren Aufmerksamkeitsverteilung durch erhöhte SA scheint SAT/ATC diese eher zu behindern.

Diesen Befund kann man im Zusammenhang mit den empirischen Untersuchungen zur unwillkürlichen Aufmerksamkeitslenkung und den daraus resultierenden Wechselkosten (u.a. Folk et al. 1992) sehen. Untersuchungen zur visuellen Suche zeigen eine automatische Bindung der Aufmerksamkeit durch plötzlich auftretende Helligkeitsunterschiede (Yantis & Hillstrom 1994) bzw. durch Blink- und Farbreize (Hüttner et al. 1995). Es ist abzuwarten, ob sich die Aufmerksamkeitsbindung durch Farbreize auch in den Blickdaten der Evaluation von SAT/ATC widerspiegelt. Insbesondere die rote Farbe, die als Markierung der Konfliktpartner verwendet wird, scheint problematisch zu sein. Als Warn- und Gefahrensignal wirkt die rote Markierung stark aufmerksamkeitsbindend (u.a. Balzert 1988). Durch eine Veränderung der Farbwahl kann der Effekt der unwillkürlichen Aufmerksamkeitslenkung möglicherweise gemindert werden.

Die erhöhte Häufigkeit überzähliger Eingriffe beim Training mit SAT/ATC und insbesondere in den Simulationen mit Unterstützung durch SAT/ATC kann im Zusammenhang mit den Interviewdaten zur Verunsicherung durch lange und unregelmäßige Systemantwortzeiten von SAT/ATC gesehen werden.

Untersuchungen zum Einfluss von Systemantwortzeiten (Voigt et al. 1995) auf die Beanspruchung und Situationseinschätzung der Benutzer zeigen neben einer Verunsicherung bei langen Systemantwortzeiten insbesondere eine negative Wirkung von variierenden Systemantwortzeiten (Hüttner et al. 1995). Variierende Systemantwortzeiten verhindern, dass sich der Benutzer auf das System einstellt und erhöhen dadurch die Unsicherheit und psychische Beanspruchung.

Abschließend wird das evaluierte SAT/ATC-System von der Versuchsgruppe unterstützend akzeptiert, aber nicht als lernförderlich beurteilt.

Ausblick

Die vorläufigen Ergebnisse der Evaluation zeigen, dass das wichtigste Problem, das in SAT/ATC gelöst werden muss, das Antwortzeitverhalten des kognitiven Modells ist. Profiling-Werkzeuge für kognitive Modelle (sind sie in Prolog, ACT-R oder SOAR implementiert) stehen jedoch nicht zur Verfügung. Es ist deshalb erforderlich, eine methodische Grundlage für das *deployment* von kognitiven Modellen, also deren Implementierung in einer Anwendung, zu entwickeln.

Erst die Auswertung der Blickbewegungsdaten wird zeigen, ob auch die Art der Aufmerksamkeitssteuerung geändert werden muss. Bei zukünftigen Entwicklungen in anderen Domänen sind auf jeden Fall Änderungen erforderlich (s.u.). Hier sollte durch begleitende formative Evaluationen die Güte der benutzten Methoden frühzeitig getestet werden.

In weiteren Evaluationen von zukünftigen SA-Trainings für den Bereich komplexer und dynamischer Mensch-Maschine-Systeme müssen methodische Überlegungen zu den Bereichen *Expertise*, *Aufgabengestaltung* und *Bewertungskriterien* einfließen. Es ist jeweils eine Abwägung zwischen wirtschaftlich machbarem und Erkenntnisnutzen zu treffen.

Eine wichtige Entscheidung zur Reichweite von Untersuchungen über solche Trainingssysteme liegt im Expertisecharakter von SA begründet: Die Untersuchung wurde an Studierenden durchgeführt, die nicht die Vorbildung und Auswahligenschaften von Fluglotsenschülern haben. Training von SA ist ein Prozess, in dem die Trainees effektive und effiziente Strategien entwickeln. Die oberflächliche Einführung und das vergleichsweise kurze Training mit dem System für die Versuchspersonen können dazu führen, dass weniger die Effektivität der Unterstützung, durch die angemessene Strategien entwickelt werden sollen, als viel mehr die Unterstützung bei einer ersten Einführung in die Aufgabe gemessen wird.

Um in dieser Situation die Anforderungen an die Studierenden zu reduzieren, wurde die Komplexität der trainierten Flugsicherungsaufgabe mit dem Ziel reduziert, die kognitiven Prozesse nicht zu verändern. Eine weitere Limitation der vorgestellten Untersuchung liegt darin, dass die Trainingsdurchläufe im Sinne eines Simulationsexperimentes aus einer Aufeinanderfolge von kurzen Szenarien bestehen. Dies ist notwendig, um Vergleichbarkeit und Kontrollierbarkeit der Bedingungen zu gewährleisten. Ein einzelnes langes Szenario hätte jedoch die realistischeren Anforderungen und würde dem anzustrebenden Trainingsbetrieb eher gleichen.

Das Konzept zum modellbasierten SA Training wird derzeit auf die Domäne Prozessführung chemischer Prozesse übertragen. Die variable Systemstruktur, die zustandsabhängige Dynamik des Systems, die zumindest abschnittsweise notwendigen kognitiven Prädiktions- und Planungsleistungen sowie die Gleichzeitigkeit von Aufgaben sprechen dafür, dass dem Konzept SA auch in dieser Domäne eine hohe Bedeutung zukommt.

Allerdings unterscheiden sich die Aufgaben der Operateure in der Prozessführungsdomäne deutlich: Typische Bedienhandlungen sind Sollwertveränderungen, Änderungen in der Nutzung der Betriebsmittel und Kontrolle der Hilfsaggregate (Johannsen 1993). Die Operateure geben Vorgaben an die Prozesse, überwachen die Funktionen und leiten in Ausnahmesituationen geeignete Maßnahmen ein (Schuler 1998). Hinzu kommt das Management von Alarmmeldungen, sowie die Beherrschung der nur teilweise automatisierten Funktionen wie An- und Abfahren.

Für die im Rahmen einer Aufgabenanalyse identifizierten Teilaufgaben „Selektion von relevanten Zustandsgrößen“, „Bewertung einzelner Zustandsgrößen“ und „Bewertung von Konstellationen“ in den Grafikbildern der Bedienstationen ist der Farbkodierungs-Ansatz aus der Flugsicherung prinzipiell übertragbar. Bei der Prozessführung mit Bildschirmen wird die Farbkodierung von Objekten jedoch bereits zur Anzeige von Systemzuständen oder Meldungen genutzt (VDI/VDE 3699 Blatt 5). Weitere Schwierigkeiten bei der Evaluation ergeben sich in der höheren Subjektivität der Leistungsbewertung. Eine eindeutige Vergabe von Labeln wie im SAT/ATC-Kontext ist hier im Allgemeinen nicht gegeben. Darüber hinaus ist zumeist eine explizite Beurteilung dynamischer Vorgänge erforderlich. Hierzu stehen den Operateuren Darstellungen des zeitlichen Verlaufs von einzelnen Prozessgrößen in Form von Kurvenbildern zur Verfügung (VDI/VDE 3699 Blatt 4). Die hier zu erbringenden kognitiven Leistungen sind derzeit Gegenstand der Forschung (Kindsmüller & Urbas 2002).

Die Entwicklung für diese Domäne sollte durch die Bewertung einzelner Fragestellungen begleitet werden, die während der Entwicklung beantwortet werden müssen: Wie ist die Aufmerksamkeit am günstigsten zu lenken, ohne die Wechselkosten zu erhöhen? Wie können bedeutungsrelevante Verläufe in den Kurvendarstellungen hervorgehoben werden? Welche zusätzliche Hilfen und Rückmeldungen müssen für das Strategienlernen vorgesehen werden? Für einen derartigen Entwicklungsprozess empfiehlt sich die Orientierung an einem parallel-iterativen Vorgehensmodell, das gezielte Evaluationschritte in allen Phasen der Systementwicklung vorsieht (Urbas & Timpe 2002).

Für die abschließende summative Evaluation, wie auch für Untersuchungen während der Entwicklung bietet sich das Vorgehen einer „Layered Evaluation“ an, bei dem mehrere abstrakte Datenverarbeitungsschritte innerhalb eines adaptiven Systems einzeln untersucht werden (Weibelzahl & Weber 2002).

Das Ziel der vorgestellten Evaluation war, die Qualität des entwickelten Trainingssystems SAT/ATC transparent zu machen. SA-Training ist schwierig zu überprüfen, weil SA ein abstraktes Konzept ist und solch ein Trainingssystem viele einzelne Fehlerquellen haben kann. Weibelzahl & Weber (2002) schlagen *Layers* für die Evaluation von adaptiven Softwaresystemen vor, von denen für das modellbasierte SA-Trainingssysteme die Trennung in *Inferenz*, *Adaptionsentscheidung* und *Interaktion* anwendbar sind. An dieser Trennung wird deutlich, dass die innere „Intelligenz“ des Trainingssystems, die kognitive Simulation, nicht als einzelner *Layer* getestet wurde, so dass insbesondere eine Bewertung dieser Einzelfunktion schwer fällt. Eine Evaluation dieses *Layers* hätte durch ein „Wizard of Oz“-Design erreicht werden können (Vergleich des vorhandenen kognitiven Modells durch Austausch mit allwissendem kognitivem Modell = erfahrenem Fluglotsen).

Diese Arbeit wird von der VolkswagenStiftung im Rahmen des Programms „Nachwuchsgruppen an Universitäten“ unterstützt.

Literatur

- Balzert, H. (1988). Einführung in die Software-Ergonomie. In *Mensch-Maschine-Kommunikation*. Berlin: de Gruyter.
- Bass, E. J. (1998). Towards an Intelligent Tutoring System for Situation Awareness Training in Complex, Dynamic Environments. In B. Goettl, H. M. Half, C. L. Redfield & J. V. Shute (Eds.), *Intelligent Tutoring Systems. Proceedings of the 4th International Conference, ITS'98. San Antonio, Texas, USA. August 1998* (pp. 26-35). Berlin: Springer.
- Chappell, A. R. & Mitchell, Ch. M. (1997). Intelligent tutoring systems for 'trained novices': Building expertise in mode management. *Proceedings of the 1997 International Symposium on Aviation Psychology, Co-*

lumbus. OH.

- Endsley, M. R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley & D. J. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. 147-174). Mahwah, NJ: Lawrence Erlbaum.
- Endsley, M. R. & Smolensky, M. W. (1998). Situation Awareness in Air Traffic Control: The Picture. In M. Smolensky & E. Stein (Ed.), *Human Factors in Air Traffic Control* (pp. 115-154). New York: Academic Press.
- Folk, C. L., Remington, R. W. & Johnston, J. C. (1992). Involuntary covert orienting is dependent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 1030-1044.
- Götte, A. W., Urbas, L. & Wozny, G. (1998). Operatorschulung in der Prozeßtechnik - Status, Trends und Anwendungen. In UMSICHT (Ed.), *Rechneranwendungen in der Verfahrenstechnik. UMSICHT-Tage 1998, Oberhausen, 15.-16.9.1998* (Vol. 7, pp. 11.1-11.11). Stuttgart: Fraunhofer IRB.
- Gopher, D. (1993). The Skill of Attention Control: Acquisition and Execution of Attention Strategies. In D. Meyer & S. Kornblum (Eds.), *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience - A Silver Jubilee* (pp. 299 - 322). Cambridge, MA: MIT Press.
- Hart, S. & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North Holland B.V.
- Hüttner, J., Wandke, H. & Rätz, A. (1995). *Benutzerfreundliche Software: Psychologisches Wissen für die ergonomische Systemgestaltung*. Berlin: Bernd-Michael Paschke Verlag.
- Johannsen, G. (1993). *Mensch-Maschine-Systeme*. Berlin: Springer.
- Kindsmüller, M. C. & Urbas, L. (2002). Der Einfluss von Modellwissen auf die Interpretation von Trenddarstellungen bei der Steuerung prozesstechnischer Anlagen. In diesem Band.
- Künstler, B. (1980). Psychische Belastung durch die Arbeitstätigkeit - Theoretisches Rahmenkonzept der Entwicklung eines Fragebogens zum Belastungserleben. *Probleme und Ergebnisse der Psychologie*, 74.
- Leuchter, S. & Jürgensohn, T. (2001). situation awareness-Training für Fluglotsenschüler. In H. Oberquelle, R. Oppermann & J. Krause (Eds.), *Mensch & Computer 2001. 1. Fachübergreifende Konferenz* (pp. 437-438). Stuttgart: Teubner.
- Loftin, R. B. & Kenney, P. J. (1995). Training the Hubble Space Telescope Flight Team. *IEEE Computer Graphics and Applications*, 15(5) 31-37.
- Mogford, R. H. (1997). Mental Models and Situation Awareness in Air Traffic Control. *International Journal of Aviation Psychology*, 7(4) 331-341.
- Niessen, C. & Eyferth, K. (2001). A model of the air traffic controller's picture. *Safety Science*, 37, 187-202.
- Schuler, H. (1998). Prozeßführung. *Chemie Ingenieur Technik*, 70, 1249-1264.
- Urbas, L. & Timpe, K.-P. (2002). Competence as a Design Goal. In H. Luczak, A. E. Çakir & G. Çakir (Eds.), *Proceedings of the 6th Int. Sci. Conf on Work with Display Units. WWDU 2002. Kongresshaus Berchtesgaden, 22-25. 5. 2002* (pp. 566-568). Berlin: ERGONOMIC Institut für Arbeits- und Sozialforschung.
- VDI/VDE 3699 Blatt 4. (1997). *Prozeßführung mit Bildschirmen: Kurven*. Berlin: Beuth.
- VDI/VDE 3699 Blatt 5. (1998). *Prozeßführung mit Bildschirmen: Meldungen*. Berlin: Beuth.
- Voigt, E., Rechenberg, U., Hallner, D., Kugler, J. & Krüskemper, G. M. (1995). Veränderungen der Befindlichkeit durch unterschiedliche Systemresponsezeiten in der Interaktion von Mensch und PC. In O. Günstürkun, R. Guski, C. Walter & A. Wohlschläger (Eds.), *Experimentelle Psychologie*. Regensburg: Roderer.
- Weibelzahl, S. & Weber, G. (2002). Advantages, Opportunities and Limits of Empirical Evaluations: Evaluation Adaptive Systems. *Künstliche Intelligenz*, 3/02, 17-20.
- Yantis, S. & Hillstrom, A. P. (1994). Stimulus-driven attentional capture: Evidence from equiluminant visual objects. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 95-107.