

- SSR Codemanagement – Ein Überblick
- Anwendung der Contextual Design Methode im Projekt FUTURE
- Speech Recognition for ATCO assistance
- ALPACA – Artificial Learning to Predict Aircraft Climb Abilities



Inhaltsverzeichnis

Übersicht der Artikel in diesem Heft	5
SSR Code Management – Ein Überblick.....	7
Antonella Seifer	
Anwendung der Contextual Design Methode im Projekt FUTURE	12
Eliana Haugg (DFS), Stephan Huber, Johanna Gramlich, Simon Mundschenk & Selina Paul (JMU Würzburg)	
Speech Recognition for ATCO assistance	19
Dr. Konrad Hagemann, Michael Slotty & Tim Albrecht	
ALPACA – Artificial Learning to Predict Aircraft Climb Abilities	28
Robin Weiß & Dr. Matthias Poppe	
Impressum	35

Die Autoren sind, soweit nicht anders gekennzeichnet, Mitarbeiterinnen und Mitarbeiter der DFS. Die Rechte an den Artikeln liegen bei den jeweiligen Autoren.

Übersicht der Artikel in diesem Heft

Antonella Seifer: SSR Code Management – Ein Überblick

URL: <http://d-nb.info/1202570682/34>, Persistent Identifier URN: urn:nbn:de:101:1-2020010915042856181125

Der Artikel beschreibt das Management des von Flugfunktranspondern zur Identifizierung von Flugzeugen abgestrahlten Codes, welcher notwendig ist, um ein Luftfahrzeug individuell zu identifizieren. Der Text bietet zunächst eine Übersicht der Funktionen der unterschiedlicher Code-Arten und beschreibt anschließend die allgemeine Nutzung von SSR Codes. Darauf aufbauend werden der Einsatz und die Verwaltung der SSR Codes in Deutschland und in Europa erklärt. Es folgt ein Aufzeigen der aktuellen Grenzen des SSR-Code-Managements. Abschließend bietet der Artikel einen Ausblick auf aktuelle Entwicklungen und zukünftige Möglichkeiten zur sicheren Identifikation von Luftfahrzeugen.

Eliana Haugg (DFS), Stephan Huber, Johanna Gramlich, Simon Mundschenk & Selina Paul (JMU Würzburg):

Anwendung der Contextual Design Methode im Projekt FUTURE

URL: <http://d-nb.info/120866185X/34>, Persistent Identifier URN: urn:nbn:de:101:1-2020042014294988342192

Im Rahmen des Projektes FUTURE (Fluglotsenunterstützung Endanflug) entwickelt das DFS-Forschungszentrum zusammen mit der Julius-Maximilians-Universität Würzburg Visionen für einen Lotsenarbeitsplatz der Zukunft. Die momentane Benutzungsoberfläche soll auf Basis aktueller Erkenntnisse und unter Verwendung der Contextual-Design-Methode nach Karen Holtzblatt und Hugh Beyer weiterentwickelt werden. Lotsen wurden in diesen nutzerzentrierten Designprozess von Anfang an mit einbezogen. In dessen Rahmen wurde zunächst eine Nutzungskontextanalyse durchgeführt und die gewonnenen Daten anschließend durch ein interdisziplinäres Team analysiert. Darauf aufbauend, wurden in mehreren Stufen Visionen für die zukünftige Gestaltung von Lotsenarbeitsplätzen entwickelt. Es entstanden drei Prototypen mit unterschiedlichen Bedienphilosophien, die durch Lotsen bewertet wurden und aktuell weiterentwickelt werden. Die Contextual-Design-Methode wird in vielen Branchen erfolgreich angewandt. Die Anwendung im Projekt FUTURE hat gezeigt, dass die Methode sehr gut auch in der Flugsicherung genutzt werden kann.

Dr. Konrad Hagemann, Michael Slotty & Tim Albrecht: Speech Recognition for ATCO assistance

URL: <http://d-nb.info/1209706601/34>, Persistent Identifier URN: urn:nbn:de:101:1-2020050712061615599669

Spracherkennung ist bereits seit vielen Jahren in der Ausbildung der DFS erfolgreich im Einsatz. Im Rahmen der SESAR Solution „Workstation, controller productivity“ wurde der weiterführenden Fragestellung nachgegangen, inwieweit Spracherkennung auch für den unmittelbar sicherheitskritischen, operativen Lotsenarbeitsplatz Unterstützung hinsichtlich einer erhöhten Produktivität bieten könnte. Zwei Prototypen wurden hierfür entwickelt und in einer umfangreichen, zweiwöchigen Realzeitsimulation zusammen mit Lotsen aus dem DFS Betrieb validiert. Die Ergebnisse zeigen, dass eine hohe und zuverlässige Erkennungsleistung des Spracherkenners an den Sprachgebrauch der Lotsen an Board absolute Grundvoraussetzung für ein funktionierendes und von den Nutzern akzeptiertes System ist. Es ergibt sich u.a. ein Weiterentwicklungsbedarf für das DFS Forschungszentrum und seine Partner hinsichtlich einer erhöhten Flexibilisierung der im Spracherkenner hinterlegten Phraseologie.

Robin Weiß & Dr. Matthias Poppe: ALPACA – Artificial Learning to Predict Aircraft Climb Abilities

URL: <https://d-nb.info/1211370046/34>, Persistent Identifier URN: urn:nbn:de:101:1-2020060215341374381064

Mit Hilfe eines künstlichen neuronalen Netzwerks wird die innerhalb von sechs Minuten erreichte Flugfläche von steigenden Luftfahrzeugen vorhergesagt. Dabei ist der Zeitparameter anpassbar und für die Vorhersage werden bereits verfügbare Mode-S-Daten genutzt. Im vorliegenden Artikel wird diese Methode, die die Vorhersage von Vertikalkonflikten deutlich verbessert detailliert beschrieben, diskutiert und mit anderen Vorhersagemethoden wie z.B. BADA verglichen.

SSR Code Management – Ein Überblick

Antonella Seifer

Einleitung

„SSR Code“, „Transpondercode“, „Squawk“, „Mode 3/A Code“... – Ja wie heißt es denn nun richtig? Und was ist es überhaupt? Und welche Modi gibt es noch, die der Transponder abstrahlen kann?

Ein Flugfunktransponder ist u.a. zur Identifizierung von Flugzeugen notwendig. In Deutschland besteht daher eine Transponderpflicht. In der DVO 923/2012 ist in Abschnitt 6 festgelegt, wer in welchem Gebiet einen Transponder mitführen und betreiben muss: „Bei allen Flügen, die in Lufträumen durchgeführt werden, die von der zuständigen Behörde zur Zone mit Funkkommunikationspflicht (Radio Mandatory Zone, RMZ) erklärt wurden, sind SSR (Secondary Surveillance Radar)-Transponder mitzuführen und zu betreiben, die in den Modi A und C oder in Modus S betrieben werden können, sofern nicht abweichende Bestimmungen eingehalten werden, die für den betreffenden Luftraum von der Flugsicherungsorganisation vorgeschrieben sind.“

Zur Identifizierung wurde ursprünglich ausschliesslich der Mode A genutzt. Im Mode A wird lediglich der Squawk übermittelt. Ein erweiterter Mode ist der Mode C, der neben dem Squawk noch die Flughöhe übermittelt. Seit 2009 nutzt die DFS als damals erste Flugsicherung zur Identifizierung zusätzlich Mode S. In diesem Mode werden Squawk, Flugfläche und eine 24-Bit-Adresse übermittelt, die jedem Flugzeug eindeutig zugeordnet werden kann.

Was genau ist der Squawk? Der Squawk ist ein vierstelliger SSR Code, bestehend aus den Ziffern 0 bis 7, der bei der Abfrage durch das Sekundärradar vom Transponder abgestrahlt wird. Er ist notwendig, um ein Luftfahrzeug individuell zu identifizieren oder einer Gruppe zuzuordnen. Denn Radarantennen erkennen, wo ein Flugzeug ist, nicht aber, welcher Flug sich dahinter verbirgt. Der Fluglotse bekäme demnach auf dem Radarschirm nur anhand eines Quadrates angezeigt, wo das Flugzeug ist, aber nicht, wer es ist, ob es sich dabei also z.B. um DLH537 oder BER032 handelt. Erst der SSR Code ermöglicht es dem Lotsen, das Flugzeug am Radarbild anhand der Luftfahrzeuglabel (siehe Abbildung 1) zu identifizieren.

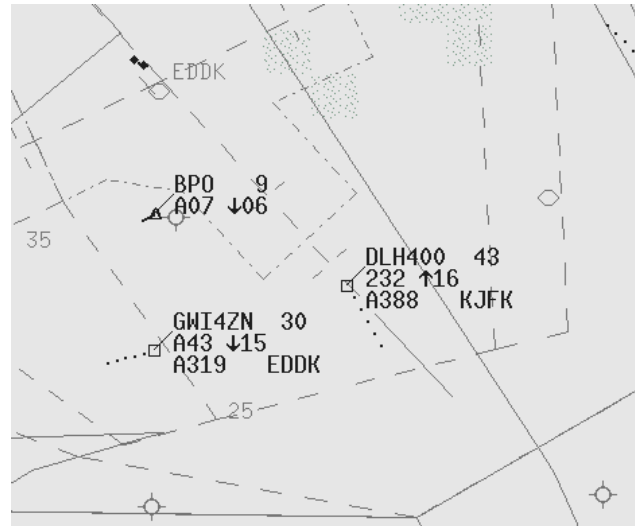


Abbildung 1: Labeldarstellung ATCAS Langen

Wer muss einen SSR Code abstrahlen?

- Jeder Flug nach Instrumentenflugregeln (IFR);
- Jeder Flug nach Sichtflugregeln (VFR) über 5000ft MSL oder 3500ft GND.

SSR Code Nutzung

Der Pilot bekommt vom Lotsen einen Squawk zugewiesen, den er in seinem Transponder einwählt. Dieser Code wird vom Flugzeug abgestrahlt und dem Lotsen auf dem Radarschirm angezeigt. Im Hintergrund erkennt das FDPS (Flight Data Processing System) den Code und korreliert das Flugziel mit dem Flugplan: Das FDPS weist dem Flugziel den entsprechenden Flugplan zu. Auf dem Radarschirm kann nach erfolgreicher Korrelation anstelle des SSR Codes das entsprechende Rufzeichen (z.B. DLH537) dargestellt werden.

VFR-Flieger nutzen Gruppencodes, die Special Use Codes (SUC). Der wohl bekannteste und am häufigsten genutzte SUC ist der Code 7000. Im ICAO European Secondary Surveillance Radar (SSR) Code Management Plan heißt es: „Code 7000 shall be used by flight crews not receiving ATS service in order to improve detection of suitably equipped aircraft in areas specified by States, unless otherwise instructed by ATS.“

Ist ein VFR-Flieger aber zum Beispiel auf der Frequenz des Fluginformationsdienstes (FIS) in Langen, wird ihm angeordnet, einen der in Langen genutzten FIS Codes abzustrahlen. Daher gibt es viele unterschiedliche

Gruppencodes für zum Beispiel lokale Flüge innerhalb von Kontrollzonen, für militärische Zwecke, für die Flugvermessung oder für das Fallschirmspringen.

IFR-Flüge bekommen ihren SSR Code vom jeweiligen DFS ATM-System zugewiesen. Dabei ist wichtig, dass jeder SSR Code nur einmal in einem Erfassungsgebiet vorkommen darf, da das System den Flugplan sonst nicht eindeutig zuweisen kann. Um das sicherzustellen, hat EUROCONTROL zusammen mit den Mitgliedstaaten die Code Allocation List (CAL) erstellt, in der festgelegt wird, welches Land/welches Gebiet welche Codes wie nutzen darf.

Hier ein paar Beispiele:

- Die Codes 3101 bis 3127 darf Deutschland zuweisen für Flüge zu allen Destinationen.
- die Codes 4350 bis 4377 dürfen nur von der FIR München zugewiesen werden für Flüge zu allen Destinationen, wenn sie über Tschechien, Österreich oder Italien ausfliegen.
- Die Codes 4701 bis 4737 dürfen bei Einflug nach Deutschland beibehalten werden, wenn der Zielflughafen in Deutschland, Belgien oder den Niederlanden ist und wenn der Flugverlauf einen Mindestabstand von 60NM von Tschechien und Polen garantiert.

Zusätzlich zu diesen fest adaptierten SSR Codes werden immer wieder Codes temporär für Sonderveranstaltungen benötigt. Bei einem Staatsbesuch zum Beispiel beantragt die Bundespolizei mehrere SSR Codes, um die Einsatzflüge eindeutig identifizieren zu können.

Aber auch bei der AERO (internationale Fachmesse für Allgemeine Luftfahrt) in Friedrichshafen, Jahresfeiern zum „Tag der Bundeswehr“ oder dem „Tag der Deutschen Einheit“ werden immer wieder SSR Codes zur temporären Nutzung beantragt. Passende Codes zu finden, ist nicht immer einfach, und hängt sehr stark von der Anzahl der beantragten Codes und dem Gebiet, in dem die Codes genutzt werden, ab.

Werden etwa Codes für ein Gebiet um Hannover beantragt, muss auf kein Nachbarland Rücksicht genommen werden. Das Gebiet ist weit genug von der Bundesgrenze entfernt, sodass die dort genutzten Codes keine angrenzenden Länder beeinflussen. Anders sieht es bei Codes für die AERO aus: in der Nähe von Friedrichshafen können nur Codes verwendet werden, die weder in Frankreich, Österreich noch in der Schweiz genutzt werden – gleichzeitig dürfen sie aber auch nicht in Langen und München genutzt werden. In diesem Fall

ist eine enge Abstimmung mit den angrenzenden Ländern und mit EUROCONTROL unumgänglich.

SSR Code Adaption in der DFS

Um diese Vorgaben einzuhalten, gibt es in den ATM-Systemen spezielle SSR Code-Adaptionen. ATCAS und iCAS gehen dabei wie folgt vor:

Bei einem Einflug:

1. schaut das System, aus welchem Gebiet der Flug in das eigene Gebiet einfliegt, und prüft, ob der bestehende SSR Code (PSSR = previous assigned SSR Code) aus dieser Region einfliegen und weiterhin genutzt werden darf. Wenn nicht, wird dem Flug ein neuer SSR Code zugewiesen (siehe SSR Code-Zuweisung).

Wenn der Code einfliegen darf

2. schaut das System, über welches Gebiet der Flug aus dem eigenen Gebiet ausfliegt und in welcher Region der Zielflughafen liegt. Es prüft, ob der PSSR unter diesen Bedingungen ausfliegen darf. Wenn der SSR Code ausfliegen darf, bleibt der Flug während des gesamten Überfluges auf seinem ursprünglichen SSR Code. Darf der SSR Code nicht ausfliegen, muss ein neuer zugewiesen werden (siehe SSR Code-Zuweisung).

SSR Code-Zuweisung:

3. Muss ein neuer Code zugewiesen werden (z.B. Start in Deutschland, PSSR kann nicht beibehalten werden) prüft das System die Flugroute, d.h. über welches Gebiet fliegt der Flug aus meinem Gebiet heraus und die Zielregion, d.h. in welcher Region landet der Flug? Anhand dieser Kriterien wird ein Codepool ausgewählt und ein freier SSR Code aus diesem Pool zugewiesen.

SSR Code Management in Europa

Grundlage für die SSR-Codeverteilung jedes EUROCONTROL-Mitgliedstaates ist der European Secondary Surveillance Radar Code Management Plan (EUR Doc 023), welcher durch die EUROCONTROL SSR Code Planning Group (SCPG) gepflegt und fortgeschrieben wird. In diesem sind die Vereinbarungen über die SSR-Codezuweisung für die Staaten und der sogenannten ATC Units der Participating Areas (PA) sowie die damit verbundenen Prinzipien und Verfahren definiert. In der European Region SSR Code Allocation List (CAL) sind die zugeteilten SSR Codes für die internationale und nationale Verwendung detailliert aufgelistet.

Die CAL koordiniert die Verwendung der 4096 Mode A Codes nach dem Prinzip der Originating Region Code Assignment Method (ORCAM [1]).

Die Verwendung der SSR Codes für internationale Flüge basiert auf der Einteilung der Europaregion in diese PAs:

- EUR-A:** Algerien, Portugal, Spanien (Canarias und Madrid FIR/UIR) und Marokko
- EUR-B:** Belgien, Frankreich, Deutschland, Irland, Luxemburg, Niederlande, Spanien (Barcelona FIR/UIR), Schweiz und Großbritannien
- EUR-C:** Dänemark, Estland, Finnland, Lettland, Norwegen und Schweden
- EUR-D:** Albanien, Österreich, Bosnien/Herzegowina, Kroatien, Zypern, Griechenland, Italien, Israel, Malta, Montenegro, Slowenien, Makedonien, Jugoslawien, Tunesien und Türkei
- EUR-E:** Weißrussland, Bulgarien, Tschechien, Ungarn, Litauen, Moldawien, Polen, Rumänien, Russland (jedoch nur Kaliningrad), Slowakei und Ukraine
- EUR-F:** Armenien, Aserbaidshan, Georgien und diverse FIRs der Russischen Föderation
- EUR-G:** Diverse FIRs der Russischen Föderation
- EUR-H:** Kasachstan, Kirgisien, Tadschikistan, Turkmenistan und Usbekistan
- EUR-I:** Diverse FIRs der Russischen Föderation
- EUR-J:** Diverse FIRs der Russischen Föderation

Codewechsel sollen nur bei Einflug in eine neue PA stattfinden. Innerhalb einer PA soll das Luftfahrzeug seinen Code während des gesamten Flugverlaufs behalten können.

SSR Code Management in Deutschland

Die Zuteilung der SSR Codes erfolgt in Deutschland durch die DFS für alle SSR Code-zuweisenden ATS-Units und Systeme der DFS sowie für die örtliche militärische Flugsicherung der Bundeswehr.

Die Verwaltung und Verwendung der zugeteilten SSR Codes durch die ATS-Units erfolgt größtenteils automatisch mittels Adaption und Verwaltung der Codes durch die verschiedenen Code- und Radardatenverarbeitungssysteme.

Mode S und Nutzung des MSCC

Die DFS hat die Nutzung des Mode S Conspicuity Code (MSCC) im März 2009 aktiviert.

Bei diesem Verfahren wird das Luftfahrzeug nicht über den SSR Code, der das Ziel mit dem Flugplan korreliert, identifiziert, sondern anhand der Mode S ID. Die Mode S ID entspricht bei IFR-Flügen dem Callsign gemäß Flugplan, bei VFR-Flügen der Registrierung. Sie wird vom Piloten direkt in den Transponder eingegeben und übermittelt. Als SSR Code muss zusätzlich der MSCC 1000 abgestrahlt werden.

Bedingung für die Nutzung des MSCC sind neben der entsprechenden Flugzeugausrüstung ein Mode S deklarerter Zielflughafen und eine Flugroute innerhalb des Mode S-Luftraums. Die Prüfung, ob der Flug für die Nutzung von Mode S geeignet ist, erfolgt über IFPS (Integrated Initial Flight Plan Processing System), ein Service des Network Manager Operations Centre (NMOC). Der Flugplan wird aufgegeben und von IFPS geprüft. Ist er für die Nutzung von Mode S berechtigt, wird im Flugplan das IFPS Flag gesetzt. Die DFS ATM-Systeme erkennen das IFPS Flag und weisen anstelle eines individuellen (diskreten) SSR Codes den MSCC 1000 zu.

Bislang wurden sieben internationale deutsche Verkehrsflughäfen als Mode S-Flughafen deklariert: Bremen, Hannover, Leipzig, Berlin-Schönefeld, Berlin-Tegel, Düsseldorf, Frankfurt/Main und München. Weitere werden sukzessive folgen. Langfristig ist geplant, den MSCC flächendeckend einzusetzen.

Grenzen des SSR Code Managements

Ein SSR Code besteht aus vier Ziffern zwischen 0 und 7. Es stehen daher insgesamt nur 4096 SSR Codes zur Verfügung, wovon die DFS wiederum nur einen Bruchteil zur Nutzung zur Verfügung hat. Das reicht nicht aus, um den steigenden Flugverkehr mit den nötigen SSR Codes zu versorgen. Allgemein sind die Codes in sog. Codepools organisiert, die ATM-systemspezifisch sind.

Aber was passiert, wenn im System keine SSR Codes mehr zur Verfügung stehen? Wird für einen Ab- oder Einflug ein neuer SSR Code benötigt, weist das ATM-System den Code aus dem passenden Codepool zu, der am längsten frei ist. Sind jedoch alle Codes des Codepools bereits „in use“ wird ein Code aus dem sogenannten „Alternate Codepool“ zugewiesen. Dieser passt jedoch nicht zwingend zur Flugroute.

Ein Beispiel: Ein Flug von München über die Schweiz nach Madrid bekommt von ATCAS München einen SSR

Code aus Codepool 11 zugewiesen. Ist kein Code in diesem Pool verfügbar, wird ein Code aus dem Alternate Codepool 10 genutzt. Die Codes in Pool 10 sind jedoch eingeschränkt und dürfen z.B. nicht in die Schweiz geschickt werden, da diese Codes dort anderweitig genutzt werden und es könnte im schlimmsten Fall zu einer Fehlkorrelation eines Ziels mit einem falschen Flugplan kommen.

Aircraft Identification (ACID) Programme

Wird die Mode S ID zur Identifizierung genutzt, muss nicht auf den Pool dieser 4096 Codes zugegriffen werden. Lediglich der SSR Code 1000 wird benötigt, die Identifizierung erfolgt direkt über die einmalige Mode S ID (s.o. Abschnitt „Mode-S und MSCC Nutzung“). Es sind somit immer genügend „Codes“ verfügbar.

Um dem steigenden Bedarf an SSR Codes gerecht zu werden, hat die EU die DVO 1206/2011 in Kraft gesetzt und EUROCONTROL das Aircraft Identification (ACID)-Programm entwickelt. Diese sieht im Artikel 4 Absatz 2 vor, dass ab 02.01.2020 jede Flugsicherungsorganisation die Fähigkeit aufweist, alle IFR-Flüge per Mode S zu identifizieren.

EUROCONTROL bietet den ANSPs im ACID-Programm zwei weitere Möglichkeiten zur Aircraft Identification als Interimslösung an:

- Die Einführung von CCAMS (Centralised Code Assignment and Management System = zentralisiertes SSR Code-Vergabesystem), oder
- die Implementierung von eORCAM (enhanced Originating Region Code Assignment Method).

CCAMS ist ein Programm des Network Managers, das das SSR Code Management für die Flugsicherungsorganisationen übernimmt, die CCAMS in ihren ATM-Systemen implementiert haben. Dazu gehören derzeit u.a. Dänemark, Schweden und Großbritannien.

eORCAM [2] ist eine Erweiterung der ORCAM-Vorgaben. SSR Codes sollen durch eORCAM zielgerichteter und flexibler genutzt werden. Bisher hat jedoch keine Flugsicherungsorganisation eORCAM implementiert.

Da Mode S die alleinige langfristige und auch mandatierte Lösung darstellt, hat sich die DFS zur Ausweitung von Mode S entschieden und die Nutzung des MSCC im März 2009 aktiviert.

Durch das ACID-Programme und besonders durch die Einführung von CCAMS können immer mehr SSR Codes bei Einflug in eine andere PA beibehalten werden. Dies wird langfristig zu einer Verschiebung oder dem Wegfall

von (einzelnen) ORCAM PA-Grenzen führen. Letztendlich bedeutet ein solcher Schritt weniger Transmissions und damit weniger Frequenzbelastung und weniger mögliche Fehlerquellen.

Aktueller Einsatz von Mode-S bei der DFS

Die DFS hat ihren Luftraum (komplett ab FL100, in vielen Bereichen tiefer bzw. bis GND) sowie aktuell acht internationale deutsche Verkehrsflughäfen als Mode S-fähig deklariert:

- alle sogenannten Major Hubs: Frankfurt/Main, München, Düsseldorf, Berlin-Tegel und Berlin-Schönefeld (somit auch BER ab 10/2020),
- sowie aufgrund der bestehenden Abdeckung auch Bremen, Hannover und Leipzig.

In den FIRs Langen, München und Bremen wird ca. jedem vierten Flieger bei Zuweisung eines neuen SSR Codes der MSCC 1000 zugewiesen. Etwa jeder dritte IFR-Flug in diesen FIRs nutzt den MSCC.

Ausblick

Die DFS wird im Zuge des Projektes MaRS (Modernisation and Replacement of Surveillance Infrastructure) alle noch nicht Mode S-fähigen Radaranlagen zu Mode S-Radaranlagen umrüsten bzw. durch ADS-B (Automatic Dependent Surveillance Broadcast) ersetzen. Eine Mode S-Abdeckung ist dann flächendeckend bis ca. 3000ft Ground gewährleistet.

Abkürzungsverzeichnis

ACID	Aircraft Identification Programme
ADS-B	Automatic Dependent Surveillance Broadcast
ANSP	Air Navigation Service Provider (dt. Flugsicherungsorganisation)
ATCAS	Air Traffic Control Automation System
ATM	Air Traffic Management
ATS	Air Traffic Service (dt. Flugverkehrskontrolldienst)
CAL	Code Allocation List
CCAMS	Centralised Code Assignment and Management System
DVO	Durchführungsverordnung
eORCAM	enhanced Originating Region Code Assignment Method

FDPS	Flight Data Processing System (dt. Flugplandatenverarbeitungssystem)
FIR	Flight Information Region (dt. Fluginformationsgebiet)
FIS	Flight Information Service (dt. Fluginformationsdienst)
GND	Ground (dt. Grund)
ICAO	International Civil Aviation Organization
iCAS	iTEC Centre Automation System
IFPS	Integrated Initial Flight Plan Processing System
IFR	Instrument Flight Rules (dt. Instrumentenflugregeln)
MaRS	Modernisation and Replacement of Surveillance Infrastructure
Mode C	Altitude Reporting Mode of Secondary Radar
Mode S	Mode Select od. Selective Co-operative Secondary Surveillance System
MSCC	Mode S Conspicuity Code
NM	Network Manager
NMOC	Network Manager Operations Centre
SUC	Special Use Code
VFR	Visual Flight Rules (dt. Sichtflugregeln)

Referenzen

- [1] ICAO (2012). European Secondary Surveillance Radar (SSR) Code Management Plan. EUR Doc 023, Amendment 1, Ed. 1.
- [2] EUROCONTROL (2011). Aircraft Identification Programme Operational Concept Document, Ed. 1.0

Anwendung der Contextual Design Methode im Projekt FUTURE

Eliana Haugg (DFS), Stephan Huber, Johanna Gramlich, Simon Mundschenk & Selina Paul (JMU Würzburg)

Einleitung

Im Rahmen des Projektes FUTURE (Fluglotsenunterstützung Endanflug) entwickelt das DFS-Forschungszentrum zusammen mit Mitarbeitern und Studenten der Julius-Maximilians-Universität Würzburg (Studiengang Mensch-Computer-Systeme) Visionen für einen möglichen Lotsenarbeitsplatz der Zukunft.

Das Projekt FUTURE wird vom Bundesministerium für Wirtschaft und Energie im 5. Luftfahrtforschungsprogramm gefördert. Es hat zum Ziel, auch unter komplexer werdenden Randbedingungen und veränderter Verkehrszusammensetzung die begrenzten Pistenkapazitäten in Deutschland effizient zu nutzen. Hierzu sollen prototypische Assistenzsysteme entwickelt werden, die den Lotsen bei der Erstellung optimaler Abstände auf dem Endanflug unterstützen [1].

Das aktuelle Human-Machine-Interface soll auf Basis aktueller Erkenntnisse zu Usability und HMI-Entwicklung nutzerzentriert weiterentwickelt und mit einer Spacing Assistenz ergänzt werden. Hierzu soll zum einen ein evolutionärer Arbeitsplatz konzipiert werden, der ohne hohen Schulungsbedarf zeitnah eingeführt werden könnte. Zum anderen soll auch ein explorativer Ansatz entworfen werden, der neuartige Anzeige- und Interaktionsmodelle aus der Forschung in den Vordergrund stellt und deshalb bewusst von aktuell verfügbaren Technologien oder Migrationsfragen losgelöst ist.

Methode

Um Visionen für einen zukünftigen Arbeitsplatz zu entwickeln, wurde die Contextual Design Methode nach Karen Holtzblatt und Hugh Beyer angewandt [2]. Die Methode wird seit Jahren erfolgreich in verschiedenen Bereichen eingesetzt. Im Jahr 2013 haben die Autoren die Methode überarbeitet, um der Art und Weise Rechnung zu tragen, in der Technologie und Digitalisierung das Leben der Menschen heute beeinflusst und verändert hat [3] [4].

Die Methode besteht im Wesentlichen aus drei Teilen. Im Rahmen einer Datenerhebung vertieft sich das Untersucherteam zunächst durch teilnehmende Beobachtungsstudien in die Welt der Nutzer und interpretiert und kommuniziert die Daten mit Hilfe von

Diagrammen und Modellen. Je nach Einsatzgebiet kann die Methode angepasst und unterschiedliche Modelle erstellt werden. Dann wird dieses Wissen genutzt, um Visionen zu erarbeiten und neue Konzepte und Produkte zu entwerfen. Schließlich werden konkrete Produkte designt und durch Nutzerfeedback immer weiter angepasst. Der vorliegende Artikel beschreibt beispielhaft den Einsatz der Methode im Projekt FUTURE. Neben der Darstellung der einzelnen Schritte und der Ergebnisse in FUTURE soll auch die Anwendbarkeit und Eignung der Contextual Design Methode im Bereich der Flugsicherung bewertet werden.

Datenerhebung

Lotsen wurden in dem nutzerzentrierten Designprozess von Anfang an mit einbezogen. Im Rahmen dieses Prozesses wurde eine Nutzungskontextanalyse durchgeführt. Diese dient u.a. dazu, detaillierte Informationen darüber zu erfassen, wie die Lotsen in ihrer normalen Arbeitsumgebung mit verschiedenen Artefakten und Kollegen interagieren, um ihre Aufgaben zu meistern. Die teilnehmenden Beobachtungen folgen dem Lehrling-Meister-Prinzip. Beobachter erfassen Informationen sowohl durch Beobachtungen des Benutzerverhaltens als auch durch Gespräche mit dem Benutzer während seiner Arbeit und versuchen, Verhalten und Strategien des Experten zu verstehen. Ein Schlüsselaspekt dieser Technik ist es, mit dem Benutzer zusammenzuarbeiten und sich von seiner Arbeit und den Problemen, auf die er stößt, leiten zu lassen. Die wichtigsten Erkenntnisse aus dieser Technik sind, zu erfahren, was die Benutzer tatsächlich tun und warum sie es so tun, sowie über erfasste Wünsche und Probleme die Bedürfnisse der Benutzer zu extrahieren.

Im Mai 2019 wurden insgesamt 8 Lotsen der Niederlassung München an den Arbeitsplätzen München und Leipzig Approach (Feeder Position) beobachtet, um ein Verständnis für die Arbeitsaufgaben und die Arbeitsumgebung zu entwickeln. Ein erfahrener Lotse stand den Beobachtern als „Dolmetscher“ zur Seite und erklärte, was gerade geschah. Dieses Vorgehen wurde gewählt, um die beobachteten Lotsen möglichst wenig bei ihrer Arbeit zu stören. Sofern es möglich war, kommentierten die beobachteten Lotsen ebenfalls während der Beobachtung ihr Vorgehen oder

beantworteten Fragen. Im Anschluss wurde mit den beobachteten Lotsen ein einstündiges Interview durchgeführt, um auch Strategien, Motivation und Bedürfnisse der Lotsen zu verstehen. Die Antworten der Lotsen wurden hierbei sowohl notiert als auch auf Band aufgenommen. Die Teilnahme an der Studie sowie die Zustimmung zur Datenerhebung und -analyse erfolgte freiwillig. Die Beobachtungen und Interviews wurden in den Schichtplan der Lotsen integriert und galten als Arbeitszeit.

Um das Bild abzurunden und auch die Schnittstellen zum Approach kennenzulernen, wurden ebenfalls vier Lotsen im Tower München beobachtet sowie Lotsen und Lufthansa Piloten während des Joint Operational Incidents Trainings (JOINT¹) befragt. Zusätzlich wurde an zwei Streckenerfahrungsflügen mit Lufthansa und Condor teilgenommen.

Konsolidierung der Daten und Erarbeitung von Visionen

Die gewonnenen Daten aus den Interviews wurden direkt im Anschluss durch ein interdisziplinäres Team aus Ingenieuren, Informatikern, Psychologen, und HCI-

Forschern in einer jeweils einstündigen Audiokonferenz analysiert und die wichtigsten Themen und Erkenntnisse erfasst. Es ist von Vorteil, die Interpretation zeitnah durchzuführen, da so die Erinnerungen an die Beobachtung und das Interview noch frisch sind. Während der Audiokonferenz wurden einzelne Beobachtungen notiert und Arbeitsmodelle erstellt, um die verschiedenen Aspekte der Arbeit zu verstehen, die für die Gestaltung des neuen Arbeitsplatzes von Bedeutung sind. Diese Modelle und Notizen einzelner Beobachtungen wurden nach abgeschlossener Datenerhebung zusammengeführt, um die Gemeinsamkeiten und Unterschiede zwischen allen beobachteten Lotsen darzustellen. Generell beschreibt das kontextbezogene Design nach Holtzblatt und Beyer [4] verschiedene *Experience*- und Arbeitsmodelle, die zur Modellierung der Arbeitsaufgaben und Details der Arbeitsumgebung verwendet werden. Im Rahmen des Projektes wurden hiervon das *Collaboration Model*, *Physical Model* und *Identity Model* als relevant betrachtet.

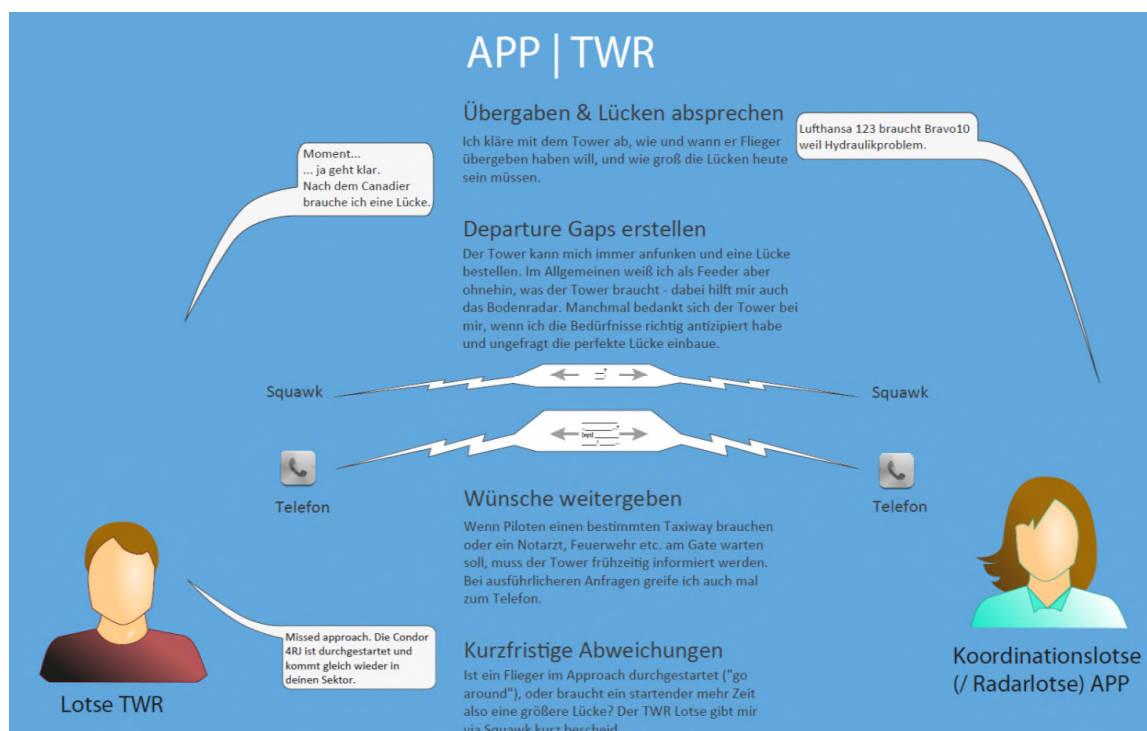


Abbildung 2: Ausschnitt aus dem Collaboration Model (

Quelle: Julius-Maximilians-Universität Würzburg, <https://www.incontextdesign.com/>).

¹ Im Rahmen des JOINT Trainings simulierten Lotsen der Deutschen Flugsicherung zusammen mit Piloten der Lufthansa

bis zur Einstellung des Programms 2019 gemeinsam Luftnotlagen.

Das *Collaboration Model* beleuchtet verschiedene Aspekte der Kommunikation mit angrenzenden Arbeitsplätzen und Akteuren, beispielsweise die Kommunikation mit Pick-ups, Towerlotsen, Wachleitern und Piloten, und zeigt sowohl Kommunikation in Standardfällen als auch Vorgehensweisen bei Notfällen auf. Das Modell geht im Sinne einer ganzheitlichen Betrachtungsweise auch auf Kommunikation außerhalb der Arbeitszeit ein, und stellt dar, inwiefern die Lotsen sich darüber hinaus mit Kollegen oder Familie über ihre Arbeit austauschen oder Kontakte pflegen. Abbildung 2 zeigt einen beispielhaften Ausschnitt aus dem erstellten *Collaboration Model*.

Das *Physical Model* (Abbildung 3) stellt die physikalische Umgebung dar, in der die Arbeitsaufgaben ausgeführt werden. Hier wurden alle Bildschirme, ihre Anordnung und alle Eingabemedien erfasst, die den beobachteten Lotsen auf der Feeder Position zur Verfügung standen.

Das *Identity Model* greift verschiedene Aspekte dessen auf, wie sich die Lotsen selbst sehen, wie sie ihre Arbeit ausführen und was sie daran mögen. Dies muss nicht auf einen einzelnen Aspekt beschränkt sein, Lotsen können sich durchaus als Vertreter verschiedener Kategorien oder Eigenschaften sehen, welche je nach Verkehrslage



Abbildung 3: *Physical Model* (
Quelle: Julius-Maximilians-Universität Würzburg).

und individueller Beanspruchung stärker ausgeprägt sind oder eher angewendet werden. Abbildung 3 zeigt anhand des „Bastlers“ einen beispielhaften Ausschnitt aus dem *Identity Model*.

Hierbei ist es wichtig zu erwähnen, dass die Modelle keinen Anspruch auf Vollständigkeit erheben, sondern



Abbildung 4: Ausschnitt aus dem *Identity Model*
(Quelle: Julius-Maximilians-Universität Würzburg, <https://www.incontextdesign.com/>).

lediglich die Ergebnisse der vor Ort gemachten Beobachtungen und Interviews widerspiegeln. Diese stellen nur einen Ausschnitt der Lotsentätigkeit und möglicher Ereignisse dar.

Um die einzelnen Beobachtungen zu konsolidieren wurde ein *Affinity Diagram* aus den Interviewnotizen erstellt. Hierzu wurden die Notizen nach der Ähnlichkeit ihres Inhalts gruppiert. Die Gruppen wurden in einem bottom-up Prozess mit farbigen Klebezetteln gekennzeichnet, wobei jede Farbe eine bestimmte Ebene in der Hierarchie darstellt. Gelbe Klebezettel wurden für einzelne Beobachtungen, blaue zur Zusammenfassung in Gruppen, pinke zur Zusammenfassung in Hauptgruppen, und schließlich grüne Klebezettel zur Zusammenfassung in Bereiche verwendet. Im vorliegenden Projekt wurden insgesamt 435 einzelne Beobachtungen notiert, die in 149 Gruppen, 41 Hauptgruppen und 8 Bereiche eingeordnet wurden. Abbildung 4 zeigt beispielhaft Notizen aus dem *Affinity Diagram* zum Bereich Kommunikation zwischen Approach und Tower.

Kommunikation App TWR
Ich antizipiere mit Hilfe des Bodenradars und der "Triebwerke-laufen-Anzeige" wann der TWR Lücken braucht und vermeide so Kommunikation über die Skquawk Box
Das Bodenradar ermöglicht mir, Lücken für den Tower zu machen und verbessert die Zusammenarbeit
Ich brauche das Bodenradar um die Bedürfnisse des Towers mit einplanen zu können
Ich schaue lieber auf das Bodenradar um die Departure-Lücken mit dem Tower zu ermöglichen
Ich schaue immer, wenn ich LFZ zum Tower schicke hoch zum Bodenradar - falls ich noch eine Speed Reduction mitgeben muss
Eine Anzeige am Radar gibt an, wie viele LFZ auf der Nord/Süd Bahn starten wollen

Abbildung 6: Ausschnitt aus dem *Affinity Diagram* (Quelle: Julius-Maximilians-Universität Würzburg).

Beyer & Holtzblatt betonen, wie wichtig es ist, das gesamte *Affinity Diagram* in ein oder zwei Sitzungen zu erstellen, anstatt kleinere Diagramme über viele Sitzungen zu erstellen. Dieses Eintauchen in die Daten über einen längeren Zeitraum hinweg hilft Teams, das breite Spektrum eines Problems schnell zu erkennen und fördert einen Paradigmenwechsel im Denken statt der Assimilation von Ideen [2].

Nach Fertigstellung des *Affinity Diagrams* wurde ein Workshop durchgeführt, bei dem ein interdisziplinäres

Team aus insgesamt 12 Ingenieuren, Informatikern, Psychologen, HCI-Forschern und ehemaligen Lotsen in einem sogenannten *Wall Walk* das Diagramm und die erstellten Modelle abliefen. Dabei konnten die Teilnehmer die Notizen top-down lesen und sich so ein vollständiges Bild der Zusammenhänge machen. Der Vorteil der Methode liegt darin, dass auch komplexe Daten erfasst und in verständliche Teile zerlegt werden können, ohne die Vielfalt der Daten oder Eindrücke zu schmälern. Während des *Wall Walks* bekamen die Teilnehmer Klebezettel, um beim Lesen entstehende Designideen zu notieren und an der Wand anzubringen.



Abbildung 5: *Wall Walk* mit *Affinity Diagrammen* und Arbeitsmodellen (Bild: Julius-Maximilians-Universität Würzburg).

Insgesamt wurden 157 Ideen notiert, die sich teilweise überschneiden. Zum Abschluss des Workshops wurde eine Liste mit sogenannten *Hot Ideas* erstellt, die die wesentlichen Aspekte zusammenfasst, um einen ganzheitlichen zukünftigen Arbeitsplatz zu gestalten. Abbildung 5 zeigt Teilnehmer beim *Wall Walk*.

Um Visionen zu erarbeiten, wie neue Konzepte und Technologien die Arbeit der Lotsen besser unterstützen können, wurden drei weitere Workshops durchgeführt. In den *Visioning Sessions* wurden ausgehend von einer oder mehreren *Hot Ideas* Szenarien entwickelt, die die Ideen zum zukünftigen Arbeitsplatz darstellen. Dabei wurde das erdachte Szenario von einer Person der Gruppe auf einem Flipchart bildlich festgehalten. Insgesamt wurden sechs Visionen entwickelt.

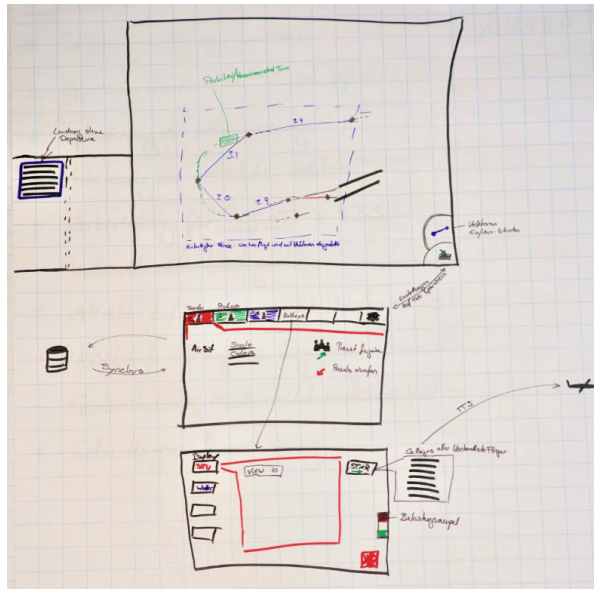


Abbildung 7: Vision „One face to ATC“ (Bild: Julius-Maximilians-Universität Würzburg).

Abbildung 6 zeigt beispielhaft eine der Visionen. Bei der Vision „One Face to ATC“ geht es darum, die relevanten Informationen in ergonomischer Weise an einem (größeren) Bildschirm zusammenzufügen, um Head-down Zeiten zu vermeiden und eine Fokussierung auf den eigenen Zuständigkeitsbereich zu ermöglichen.

Konzepterstellung

Aus den Visionen sind drei einfache Klickprototypen entstanden, die mit unterschiedlichen Bedienelementen nutzbar sind. Die Prototypen wurden zunächst bewusst einfach gehalten, um nicht zu viel Arbeitszeit für Details aufzuwenden, die gegebenenfalls nicht weiterverfolgt werden.

Die Konzepte beinhalten einen Prototyp, der mit der Maus bedient wird, einen weiteren der mit dem kommerziellen Microsoft Tool „Surface Dial“ und per Touchscreen bedient wird (s. Abbildung 8), diese Konzepte könnten zeitnah und ohne hohen Schulungsaufwand umgesetzt werden. Außerdem wurde ein explorativer Ansatz mit vier Tangible Entwürfen aus dem 3D Drucker entwickelt (s. Abbildung 9 links).



Abbildung 8: Prototyp Microsoft Surface Dial (Bild: Julius-Maximilians Universität Würzburg).

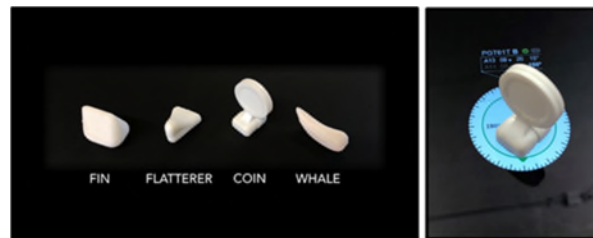


Abbildung 9: Prototypische Tangibles (links), Möglichkeit zur Heading Eingabe (rechts, Bild: Julius-Maximilians-Universität Würzburg).

Das Label der Maus und Dial Prototypen beinhaltete in der ersten Zeile Callsign und Wake Turbulence Kategorie, sowie Icons zur Identifizierung eines LFZ (grüner Haken), Freigabe des Localisers und des ILS. Diese wurden in den verschiedenen Prototypen entweder direkt per Mausklick (Togglebutton) oder über das Dial bedient (Abbildung 10 und Abbildung 11). In der zweiten Zeile standen aktuelle Werte für Flight Level/ Altitude, Rate (mit Pfeil entsprechend Steig/Sinkrate), Speed und Heading. In der Zeile darunter Vorschläge für nächste Freigaben, die vom Lotsen angenommen oder zuvor geändert werden konnten. Freigegebene Vorschläge wurden unterstrichen dargestellt. Die Vorschläge sollen zukünftig basierend auf Daten zur für die Position üblichen Freigaben mit Hilfe künstlicher Intelligenz generiert werden. Durch Bewegung der Maus über das Callsign konnten je nach Prototyp in einem erweiterten Label Informationen zur Landebahn, Flugzeugtyp, Wetterinformationen und QNH angezeigt werden.



Abbildung 10: Label des Prototypen mit Maus Bedienung (Bild: Julius-Maximilians-Universität Würzburg).

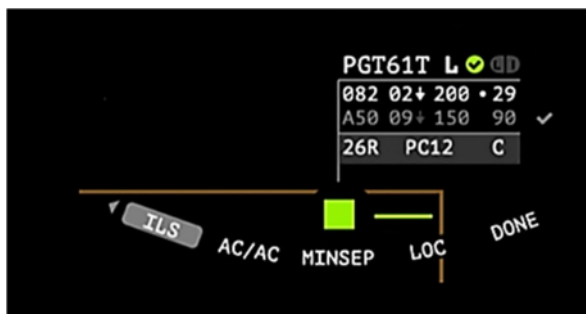


Abbildung 11: Label des Prototypen mit Dial Bedienung und erweitertem Label (Bild: Julius-Maximilians-Universität Würzburg).

Die Prototypen boten beispielsweise auch die Möglichkeit, schnell Abstände zwischen LFZ anzuzeigen, Mindestabstände einzublenden, oder Mode S Daten auszulesen.

Die Interaktion mit den auf dem Radarbildschirm positionierten Tangibles orientierte sich an vier grundlegenden Interaktionen, die während der Beobachtung identifiziert wurden:

1. Identifizieren eines LFZ durch direktes Berühren mit dem Tangible,
2. Heading ändern durch Drehung des Tangibles in die neue Richtung,
3. Neigen des Tangibles nach vorne, um die Geschwindigkeit zu erhöhen, und nach hinten, um die Geschwindigkeit zu verringern,
4. Veränderung des Flight Levels/Altitude durch Andrücken des Tangibles auf den Bildschirm (Sinken) und zu sich Ziehen des Tangibles (Steigen).

Die Prototypen wurden zunächst am 13.12.2019 dreizehn Münchner Approach Lotsen vorgestellt. Die Lotsen waren sehr offen und interessiert gegenüber den vorgestellten Konzepten und gaben zahlreiche Vorschläge zur Ergänzung und Verbesserung der Inhalte, Interaktionen und Bedienelemente. Anhand des

Feedbacks wurden die Prototypen überarbeitet und am 10.01.2020 weiteren dreizehn Münchner Approach Lotsen vorgestellt, die noch nicht am ersten Termin teilgenommen hatten. Die Ergebnisse des zweiten Termins werden aktuell noch ausgewertet. Aus dem gesamten bisherigen Feedback lassen sich aber bereits wesentliche Bedürfnisse der Lotsen ableiten:

- Minimalistische Benutzungsoberflächen,
- Möglichkeit zur Individualisierung von Anzeigen entsprechend der Arbeitsposition und individueller Arbeitsweisen,
- Effiziente und funktionale Interaktionen,
- Implementierung verschiedener Interaktionsmöglichkeiten, um ein Ergebnis zu erreichen,
- Bewusste Änderung und Bestätigung/Verwerfen von Werten,
- Visuelle Unterstützung um die Aufmerksamkeit auf relevante Daten oder Änderungen zu lenken.

Im Anschluss an die Auswertung sollen die Prototypen ein weiteres Mal aktualisiert und den Lotsen ein letztes Mal vorgestellt werden. Detaillierte Ergebnisse werden nach Abschluss gesondert veröffentlicht.

Fazit

Die Contextual Design Methode wird in vielen Branchen erfolgreich verwendet, beispielsweise im Bereich der Gestaltung von Webseiten, Apps, medizinischer Geräte, Unterhaltungselektronik oder Automobilelektronik. Die Anwendung im Projekt FUTURE hat gezeigt, dass die Methode sehr gut mit angemessenem Aufwand in der Flugsicherung anwendbar ist. Nach Holtzblatt und Beyer [4] verändern neue Technologien üblicherweise nur die Art, wie Aktivitäten ausgeführt werden, nicht die Aufgabe selbst. Demzufolge kann ein Unternehmen die erhobenen Daten etwa 3-5 Jahre lang nutzen, was den Aufwand ebenfalls rechtfertigt.

Dabei ist die Contextual Design Methode weit mehr als eine reine Tätigkeitsanalyse, wie sie bisher in der Flugsicherung immer wieder angewendet wird. Contextual Design will nicht nur Tätigkeiten protokollieren und analysieren, sondern auch die Strategien, Work-Arounds, Motivationen und Emotionen der Nutzer verstehen. Dies dient dem Ansatz, nur Komponenten oder Produkte zu entwickeln, die der Nutzer wirklich braucht. Der Einsatz von neuen Technologien soll sich an den tatsächlichen Aufgaben der Nutzer orientieren, Technologien sollten nicht nur implementiert werden, weil sie neuartig sind oder zur Verfügung stehen.

Ebenso sollte althergebrachtes überdacht werden: was ist nützlich, was überholt, welche Bedürfnisse haben die Nutzer? Neue Generationen von Nutzern, die mit digitalen Medien aufgewachsen sind, haben einen hohen Anspruch an einen Arbeitsplatz. Dieser soll sie idealerweise sowohl bei der Durchführung ihrer Aufgaben individuell und optimal unterstützen, ein intuitives, ansprechendes Design bieten, als auch Spaß machen.

Abkürzungsverzeichnis

APP	Approach
ATC	Air Traffic Control
FUTURE	Fluglotsenunterstützung Endanflug
HCI	Human-Computer-Interaction
HMI	Human Machine Interface
ILS	Instrument Landing System
JOINT	Joint Operational Incidents Training
JMU	Julius-Maximilians-Universität
LFZ	Luftfahrzeug
QNH	Höhenkorrekturwert (bezogen auf NN)

Referenzen

- [1] DFS Deutsche Flugsicherung GmbH (2017). Luftfahrtforschungsprogramm 2018-2022. Fluglotsen-Unterstützung Endanflug. Gesamtvorhabensbeschreibung FUTURE. Version 1.0.
- [2] Beyer, H. & Holtzblatt, K. (1997). Contextual Design: Defining Customer-Centered Systems. San Francisco: Morgan Kaufmann.
- [3] Holtzblatt, K. & Beyer, H. (2015). Contextual Design Evolved. Morgan & Claypool Publishers
- [4] Holtzblatt, K. & Beyer, H. (2016). Contextual Design: Design for Life: Morgan Kaufmann.

Speech Recognition for ATCO assistance

Dr. Konrad Hagemann, Michael Slotty & Tim Albrecht

Introduction

In January 2020 together with the latest update of an office software package, speech recognition became part of the standard computer equipment for DFS employees. Since this installation, every employee could use it to dictate e.g. emails just by simple button press. However, when walking down the office floors, the usage still seems to be very limited. Maybe dictating emails to a notebook appears to be too exotic for most people? It is probably similar to the confusion that comes up when somebody e.g. in the train speaks into a hidden headset: It just looks kind of odd watching somebody talking to air leaving it unclear if the one really has a conversation partner. However, the situation changes with a smartphone in hand or a visible headset applied and maybe in a more private setting than a train or office. As a matter of fact, high performing speech recognition software such as Siri, google or Alexa earn more and more followers today. Especially people who want to avoid time consuming typing for short messages or emails with large parts of standard phraseology ("Dear ladies and sirs...", "With kind regards...", etc.) increasingly learn to profit from its advantages. Together with an easy to use (often touch-based) error correction system it may soon become even more popular in frequent text communication of everyday life or add to a higher productivity at work.

But if it is already such a widespread technology why is it that difficult to use speech recognition in ATC operations? The answer is fairly easy if one tries a simple experiment by just giving any standard ATC clearance such as *"lufthansa two three two zero hello identified descend flight level eight zero at two thousand feet per minute or greater"* into any commonly available speech recognition system, e.g. a smartphone (Helmke, 2019). Currently, the recognition performance to such input is very bad. In ATC phraseology, special vocabulary is used together with a grammar that the system has not originally been developed for. To provide a system with such a specific understanding of ATC language is just not trivial and takes many years of development including many valid ATC speech data samples.

This article reports about latest R&D work at DFS pursued from 2016 – 2019 in the SESAR 2020 wave 1 project PJ.16-04 "workstation, controller productivity" (for project homepage see SESAR 2019a). Within the

project, DFS conducted a validation exercise (V2, EUROCONTROL, 2010) in October 2018 in Langen to assess the potential benefits of a speech recognition system that is used as an assistance system for operations. While this article mostly focusses on the DFS works, the results of all five exercises conducted by the various European project partners have altogether been reported elsewhere (see SESAR, 2019b).

Background

Early R&D work on voice recognition prototypes at Deutsche Flugsicherung GmbH (DFS) dates back to 1994 (Bürvenich, 1994). Based on the experiences, DFS consecutively ran real time simulations in order to integrate voice recognition into an ATC operational environment and to demonstrate potential benefits for operations (Bierwagen & Vielhauer, 2000).

In 2010, DFS decided to start the use of Voice Recognition and Response (VRR) (cited after Buxbaum, Helmke, Rühl & Slotty, 2015, p. 24; see also UFA, 2011) in the area of ATC training. In this case, VRR is acting as a simulation pilot. The VRR product ATVoice® is fully integrated into the UFA's radar/centre simulator ATCoach®. Based on a definite grammar, it recognizes transmitted clearances and converts it to text, provides the pilot's readback and feeds the simulator with the given commands for the related aircraft. New phraseology can be added as new grammar. DFS uses VRR already for simulations at the ACC in Munich and is in transition to other centres. The same simulator system and voice recognition technology is also in use at different air navigation service providers (ANSPs) (e.g. NAV CANADA), Collegiate Training Initiative Schools (CTIS) in the US (Vaughn College of Aeronautics and Technology in Flushing, NY, LeTourneau University, Texas, University of North Dakota, University of Alaska at Anchorage, Texas State Technical College at Waco, Community College of Baltimore County, etc.) and US military organizations (US Army, US Army Air Traffic Schoolhouse at FT. Rucker, Alabama, Army National Guard Airfield Operations Battalion at Ft. Bragg etc., see UFA, 2020a).

During the iPort Project (Konopka, 2013), DFS used Voice Recognition and Response for the validation in the Tower environment. Here, VRR was acting as a

simulation pilot similar to the ATC training mentioned above.

More than a decade ago, DFS System house started to enhance the T-FDPS together with Tower Düsseldorf to react on clearances given by the controller using the recognition engine. For the first phase DFS focused on clearances like landing clearances. The goal was to reduce the head down times for the controller thus to ensure safety and to increase controllers' productivity. In 2016, an internal study was performed using spoken controller clearances and the VRR system to feed the paperless strip system (PSS). Clearances like level changes, heading changes and rates were recognized by the system and the respective values were highlighted in the HMI leaving the controllers with just confirming the preselected values, i.e. executing the command.

Additional reports on DFS activities on voice recognition and response development can also be found in Slotty & Rühl (2012), Aerojournal (2013), and UFA (2017). Most recently, DFS speech recognition system provider UFA also published a press release on its SESAR 2020 activities together with DFS (UFA, 2020b) which provides a quick overview on the activities that are reported in more detail in this article.

DFS SESAR 2020 Speech Recognition ATCO Assistance Concept

Within the SESAR solution PJ.16-04, DFS developed two speech recognition-based prototypes to assist ATCOs with their Air Traffic Control task. The development was performed by an inhouse development team of the ATM Simulator Centre and based on the controller display application used for the upper airspace control training of the DFS Academy.

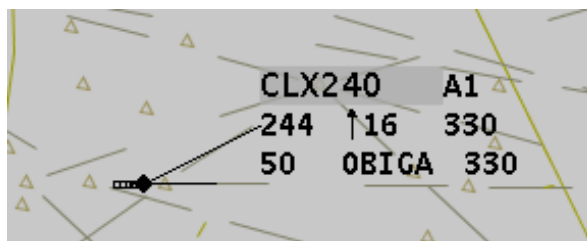


Figure 1: Target Location Assistance (TLA): The addressed callsign is highlighted with a 60% transparent grey background and a bigger font of +2px

A Target Location Assistance (TLA) was designed to enhance situational awareness by highlighting each callsign on the radar screen which had been previously addressed by the ATCO. Figure 1 shows the highlighting

which was performed within a 1-second time limit after the utterance of the callsign without the necessity of the controller clearance to be completed.

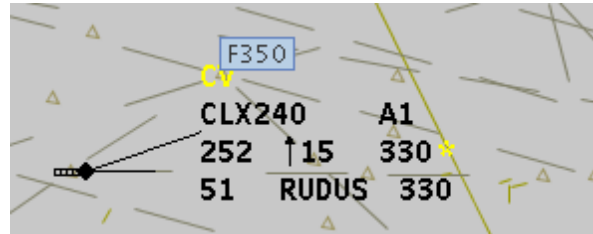


Figure 2: Controller Clearance Verification (CCV or CV): warnings are indicated via yellow letters in line 0 of the label and a yellow asterisk next to the mismatching value. A mouseover (blue box) shows the values (here flight level) recognized by the recognition system.

Controller Clearance Verification (CCV) or Clearance Verification (CV) provided a yellow warning in the upper line of the target label together with a yellow asterisk at the respective value each time a mismatch between a spoken command and a system input could be detected by the speech recognition system. Figure 2 shows an example for a CV warning. In general, the warnings were limited to detect mismatches of clearances of Flight Level, Speed, Rate and Direct. In addition, the warning also popped up after a period of 8 seconds if the corresponding system input to the recognition was missing or was still not finished.

Methods

Participants

Eight male Air Traffic Controllers (mean age = 32, range = 25-43 years) holding an active licence (mean = 8 years, range = 3-18 years) participated in groups of four per week. Seven controllers worked at Munich Approach. Due to limited availability of personnel for both weeks, one controller was recruited from Frankfurt Approach and participated only in week 2 of the simulation in the role of a feeder and only on the first two days.

Technical and Operational Set-Up

The human-in-the-loop real time simulation took part in the DFS Research Simulator (Advanced Function Simulator, AFS) in Langen. The overall validation platform consisted of the AFS simulator, the speech recognition engines and the updated emulated controller display of iCAS. Four positions of Munich Approach (for airspace see Figure 3) were simulated: North Low Executive Controller DMNLB (Pickup), South Low Executive Controller DMSLB (Pickup), Director Executive Controller



Figure 3: Munich Approach airspace.

(North) DMAT1 (Feeder) and Director Executive Controller (South) DMAT2 (Feeder). During three runs where one controller was missing for the full exercise setup the two Feeder positions DMAT1 and DMAT2 were combined to DMAT, which is also commonly used at the control center Munich during low to medium traffic.

UFA's Voice Recognition ATVoice system (UFA, 2018) was installed on notebooks that were hooked on to the simulator and placed next to each controller working position. In the Standalone Edition, the Voice Recognition support is connected to a Voice Over IP (VOIP) Communication System to provide recognition results for further processing. The system incorporates the BBN HARK AVOKE speech recognition system, a high accuracy recognition engine that is suited to ATCoach and other simulation control applications. AVOKE is a speaker independent technology, requiring no voice training or templates and is continuous, allowing the user to speak at a very rapid rate, as controllers tend to do. The ATVoice software system can operate directly with the DFS SINATRA, a Voice Over IP voice communications system (VCS).

Experimental Design

Table 1: Experimental Design (cells show number of runs per conditions).

		RECOGNITION			SUM
		off	on	CCV	
Traffic	Medium	2	4		
	High	2	4	2	
	Training	2		2	18

A two-factorial design with the main factors RECOGNITION (on/off) and TRAFFIC (medium/high) has been used (see Table 1). In addition, there were four runs



Figure 4: pseudo-randomized CV events were triggered auditorily and visually via an auxiliary screen. Stimulus presentation realized with Matlab Psychophysics Toolbox (Psychtoolbox-3, 2020).

in a so-called CCV condition that took place on the first and last day of each test week. These runs were designed to provoke a guaranteed minimum amount of CCV events (realized by visual-auditory prompts as shown in Figure 4).

As CCV was designed as a safety net feature the additional prompts were to ensure that controllers have had enough time to experience this feature during the simulation. Each time a trigger was presented ATCOs were asked to provoke any type of CCV warning of their own choice. Finally, two training runs were included at the beginning of each test week.

Schedule

The simulation took part Monday to Wednesday over two weeks in fall 2018 (22.10.-24.10. & 29.-31.10.2018) with three runs per day. Overall 18 runs were conducted. Each day started with a briefing and consisted of three runs with two runs in the morning and one in the afternoon. Two training runs (first without and second with speech recognition active) were applied at the beginning of each test week.

Data Acquisition

Workload and Situational Awareness were assessed by using online questionnaires that appeared every 5 minutes on a separate touch screen to ask the ATCOs for an instantaneous rating during the run. In addition, postrun standard questionnaires were used to evaluate workload (NASA-TLX), situational awareness (SASHA) and acceptance (CARS). Recognition rate, general operational aspects as well as some control variables related to training, simulation realism and simulator performance were assessed via a tailor-made postrun questionnaire. In debriefings, the runs as well as the general operational usability of the speech recognition

features have been reflected. Objective data were collected via simulator recordings to assess traffic load and nuisance alerts (the latter in week 2 only: ATCOs were briefed to press a button whenever they had the impression of an unnecessary warning, e.g. based on a misrecognition).

Finally, all ATCO commands spoken during the runs were recorded together with the logfiles of the speech recognizer for later offline analysis of the word recognition rate.

Data Analysis of Recognition Rate

The correct recognition of spoken utterances is an essential precondition for the validation of any application based on this technology. If the recognition rate is too low this may confound any other measurement in a validation. Therefore, a special emphasis was put on the analysis of speech recognition and several approaches were taken for its analysis.

First, right after each run, ATCOs were asked via questionnaire to subjectively estimate the recognition rate between 0-100%.

Second, a fast look at parts of the speech recognition data could be provided directly for each debriefing. To provide this, predefined keywords (callsign, flight level, speed, rate and direct) in the speech recognizer logfiles of a run were automatically compared with the corresponding commands that were documented in the logfiles recorded from the simulation pilots' inputs (reference).

Third, a more in-depth manual analysis of the predefined keywords (callsign, flight level, speed, rate and direct) was done offline for a sample part of the runs and participants. For this analysis 22 logfiles of the recognizer were each compared with all the respective audio files by a human listener and assessed for correctness.

Finally, an extensive offline analysis was performed in which each of the 58 runs (ca. 200-450 utterances per run with ca. 14,500 audio files) were validated and the recognition rate was evaluated. For the latter analysis, the Simulator Tag Accuracy was calculated, i.e. the machine commands for the simulator that were recognized (hypothesis) were compared to the machine commands derived from the transcriptions validated by a human expert (reference). Moreover, in this analysis it could be automatically differentiated between "in grammar" and "out of grammar" phrases, i.e. sentences that lied within or without the predefined capabilities of the recognizer.

Results

Recognition rate

During the debriefings, ATCOs reported that they considered the overall recognition rate as being too low for an operational use. When asked in more detail, some ATCOs estimated an error occurrence of roughly 1-5 cases per hour to be still acceptable in their safety critical environment provided that this will not increase the controllers' workload. This impression was also reflected by a mean recognition rate of 79.73% in the postrun questionnaires and the discussions from the debriefings. However, when looking at the objective data, the view appears to be a bit less pessimistic (see Figure 5 for an overview).

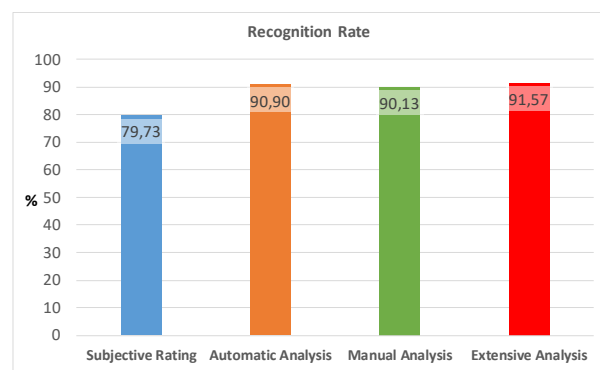


Figure 5: Recognition rate derived from subjective and objective data and analysed following various approaches.

The automatic keyword analysis revealed 90.90% correct recognitions, which was just above the minimum threshold requirement of 90% correct recognitions defined in the project. For the debriefings, this scripted analysis helped to clarify some of the limitations of the recognizer (e.g. systematic recognition problems with certain digits or waypoints with certain voices etc.) and could be discussed among the participants. The offline, manual keyword analysis over a sample of 22 runs (10,117 events) confirmed the results of the automatic analysis and revealed an almost similar overall mean recognition rate of 90.13% (min: 55.74%, max: 95.56%). Finally, the extensive Simulator Tag Accuracy analysis of all runs revealed 91.57% for in grammar phrases.

However, 31.9% of the phrases contained segments that were out of grammar (OOG), which led to an a priori exclusion of the whole phrase from the calculation of the Accuracy Score. The most common uses of OOG Phraseology were listed separately together with their frequencies of occurrence. Besides general error sources (PTT clipping off the end of the phrase, stutters or

tripping over words and use of ‘uh’ or ‘um’) overall 21 OOG categories were identified.

When looking at the CCV nuisance alerts over four solution runs, 164 (63.32%) out of 259 warnings were clicked, i.e. rated as being based on a misrecognition of the speech recognizer. This exemplary test stands in line with the overall feedback from the ATCOs in the debriefings that describes the recognition rate as being overall unsatisfactory.

Control variables

Answers to control items in the postrun questionnaire concerning the amount of training, realism and system performance suggest a high validity for the validation. Concerning the manipulation of taskload, Figure 6 and Figure 7 show the average traffic for the two scenarios per role.

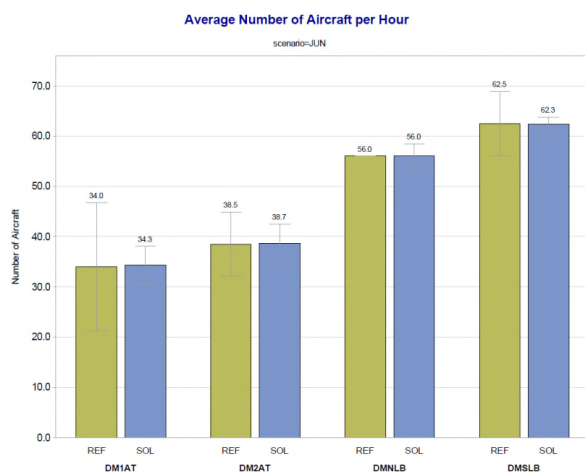


Figure 6: average number of aircraft per hour in high traffic scenario (error bars represent standard deviation).

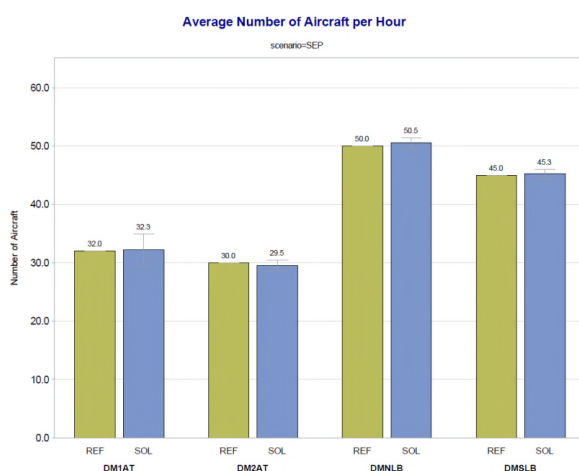


Figure 7: average number of aircraft per hour in medium traffic scenario (error bars represent standard deviation).

As can be seen by comparing the two figures, the June scenario provided higher traffic than the September scenario for all positions, with the Pick-Up South showing the greatest variation in traffic across scenarios. Neither situational awareness nor workload show any indication of differences between medium and high traffic scenarios. Therefore, the data presented in the following have been aggregated over both scenarios.

Situational Awareness

As can be seen in Figure 8, Situational Awareness is on an acceptable high level with merely any differences across experimental conditions.

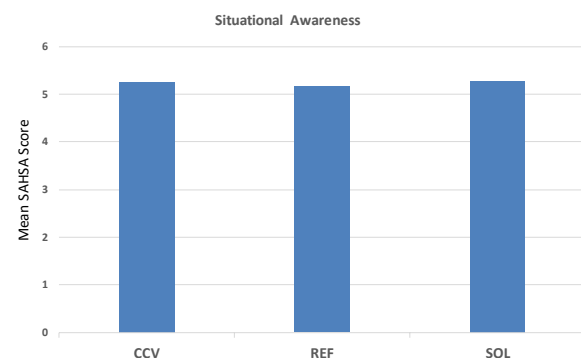


Figure 8: Situational Awareness across experimental conditions (SASHA Scores).

The values were averaged across roles as Feeder and Pick-up show hardly any variation in mean values. These results are further confirmed by the data from the online questionnaires and by the feedback obtained during the debriefings.

Workload

The workload scores (see Figure 9) show acceptable medium levels with hardly any differences across experimental conditions.

Similar to the analysis of situational awareness, the values were averaged across roles as Feeder and Pick-up show hardly any differences in mean values. Controllers were also asked to rate their peak workload in each run. Results show a similar pattern with still acceptable workload levels. These data are confirmed by the ISA scores which reflect a “relaxed” to “comfortable” level on the scale and with no significant differences across all conditions and roles.

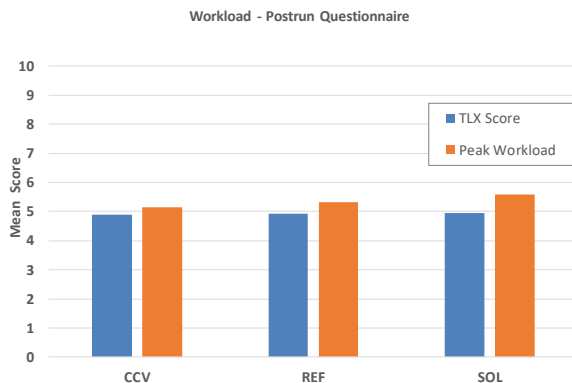


Figure 9: Workload across experimental conditions (NASA-TLX and Peak Workload Score).

Acceptance

CARS acceptance ratings for both speech-based types of assistance (TLA and CCV) are shown in Figure 10. They reveal similar mean values across conditions on a level that is labelled with “Minor but Annoying Deficiencies. Desired performance requires moderate controller compensation” (level 7).

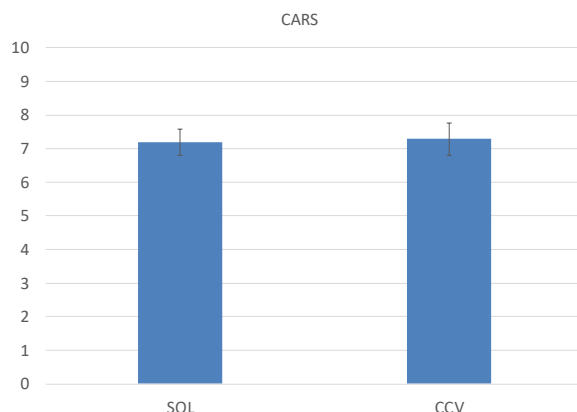


Figure 10: Acceptance across experimental conditions (CARS Scores range from 1: “Improvement mandatory. Safe operation could not be maintained” to 10: “Deficiencies are rare. System is acceptable and controller doesn’t have to compensate to achieve desired performance”, error bars represent standard error).

When asking more specifically for the initially assumed benefits of the Target Location Assistance these were not confirmed by the participants. Although the TLA display seems to be well designed (i.e. not too distracting, not too dominating the pace of work and with sufficient timeliness) the assumed benefits of this assistance with respect to faster identification of flights or improved communication with pilots cannot be agreed on. In the debriefings, the ATCOs said that although the

highlighting of call signs does no harm, it does also not represent an improvement for the operational environment.

For CCV warnings, on the one hand the feedback indicates a contribution to better detect input errors, on the other hand a significant enhancement for safety or general improvement to the operational environment cannot be seen. In the debriefings, again the false alarm rate of the speech recognition system is considered as being too high.

Discussion

The results reveal no clear benefit for using target location assistance or controller clearance verification in operations. No impact, neither on Situational Awareness nor on Workload can be observed.

Remarkably, the experimental manipulation of the factor TRAFFIC does not show any observable impact on workload. A possible interpretation is, that this may indicate a floor effect, i.e. the overall amount of traffic in both scenarios might have been too low (max. “comfortable”) so that an assistance to handle high traffic was not effectively needed. The situation could have been just easily handled with the help of the traditional setup without any assistance. Although, the feedback on “realism of the simulation” seems to speak against such an interpretation, it could still be that the assistance features only reveal their full potential under conditions of high taskload.

However, a more important impact for the validation results might have come from the recognition rate that the participating ATCOs considered as too low. Together with the experience of a high rate of nuisance alerts, these two artefacts represent confounding variables for the validation, especially for the acceptance ratings. A sufficiently high enough recognition rate, i.e. above 90%-95% (the ideal boundary value is still subject of discussion) is a precondition for any assistance based on this technology. Positive aspects (e.g. a good HMI design or high timeliness of displays) are irrelevant if the high number of misrecognitions or false alerts lead to annoying corrective actions or distraction from the traffic, especially in the case of CCV warnings.

The more extensive offline analysis revealed that in almost 1/3 of the phrases partially an OOG expression was used by the ATCOs in the simulation, i.e. phraseology that deviates from the standard as defined by DFS in the Manual of Operations Air Traffic Services (see e.g. DFS, 2019) which has been derived from the ICAO (2016) definitions. Having its origins in simulator

training, the available grammar of the speech recognition system at DFS has been strictly derived from these definitions. In fact, due to its criticality for safety, sticking to the standard phraseology is absolutely encouraged. “Pilots and Controllers might use non-standard phraseology with good intentions; however, the use of standard ICAO phraseology helps to minimize the potential for misunderstanding.” (ICAO cited after DFS, 2019).

However, humans are just different from machines in a way that they usually tend to find their own individual style and create variations in performing their work. Following this, it becomes clear that if used as an assistance system in operations, the preconditions for a speech recognition system change. Taking such a system as a disciplinary tool (which may be very reasonable for a use in training) most likely “overloads” the reasonable set of requirements for such a system in operations. Having such a system will probably never achieve high acceptance by its users but would rather be regarded as a supervisory tool introduced by the employer that tries to parent its users and uses misrecognitions as kind of a punishment for wrong behaviour. For example, couldn't it maybe be more elegant if a speech recognition system would process all utterances but would provide the possibility for an optional, individual offline safety check of used grammar? Controllers could use this as an opportunity for a confidential and voluntary self-assessment.

From looking at the OOG data, it seems that small variations in phraseology (e.g. adding a “Servus” at the beginning of a transmission) are easily rejected even if they still seem to lie within reasonable limits. To achieve high levels of satisfaction, effectiveness, and efficiency (i.e. usability, see EN ISO 9241-11:2018) systems should be designed according to the users’ needs based on their task goals. Therefore, OOG phraseologies that reflect common practices need to be carefully assessed and then the grammar of the speech recognition system needs to be expanded in order to achieve a higher level of flexibility and adaptability to the users’ tasks. This approach follows the understanding that a speech recognition should never be confused as a replacement for regular safety briefings and refresher trainings for safe use of phraseology (a recognizer transcript could nevertheless be considered helpful for such briefings though).

After all, looking back at the analysis, following multiple approaches to analyse the recognition rate turned out as a clear advantage and is recommended for any subsequent studies. Besides the subjective

assessments (questionnaires and debriefings), with this approach it had been possible to provide some directly available objective data as input at the beginning of the debriefings. For the discussions with ATCOs the data helped to identify some subjective bias (the subjective assessment revealed much lower ratings for recognition rate) and to identify potential reasons. Moreover, highly correlated results from the objective analysis methods speak for the reliability and validity of the quick analysis based on sample parts of the whole data. As an extensive offline analysis means many days of transcript validation by listening to all the audio recordings, the quick analysis is a valuable addition to the existing repository of available analysis instruments at DFS.

Outlook

The results represent another step forward for more R&D work still to come at DFS in the future. In 2020, Speech recognition technology is (once more) listed among the top innovation topics for the years 2020-2024 (DFS, 2019, p. 5-240). Almost 10 years ago, DFS developed a veritable cornucopia of potentially valid use cases (Buxbaum, 2012), two of which have now been addressed by DFS in SESAR project PJ.16-04. The collected data from the validation help to further understand the spectrum of challenges that are yet to overcome. Despite its successful application in the training environment, one must admit that bringing speech recognition into ATC operations is still more at the beginning than close to the board.

In SESAR project PJ.16-04 a selection of use cases has been addressed by partners across Europe. Besides the different results from five validations, it can be seen as one of the most significant achievements of the project that speech recognition technology is now officially anchored as a part of the European ATM Architecture (EATMA) framework. Consequently, various ANSP and industry partners can now continue their R&D work within the SESAR 2020 project PJ.10 in wave 2 with ambitious goals such as reaching technology readiness level 6 (“ready for deployment”) by the end of the project. Although not involved with an own technical contribution in wave 2, DFS is very much involved in this endeavour by taking the lead at overall project level.

Besides this engagement, DFS-internal R&D work focusses on reaching a significant improvement in the preconditions for the application of speech recognition: Most importantly, the number of misrecognitions needs to be improved to increase user acceptance. In order to achieve that, much potential lies in the further in-depth analysis of the recorded PJ.16-04 OOG data and that led

to an automatic rejection of almost 1/3 of the data. As mentioned earlier, the currently available grammar of the speech recognizer needs to be expanded to allow a higher flexibility that represents the natural environment of the ATCO users on board. Approximately 9900 OOG phrases are available and are consecutively analysed to further develop the capabilities of the system.

Finally, within the SESAR project PJ.16-04, the project partners jointly developed a common ontology to standardize transcription output. This ontology has been designed as an enabler for the exchange of speech data and it represents an important step to be able to build a large common data base for ATC like they exist for other domains (see Helmke et al., 2018). Powerful state of the art speech recognizer engines that include artificial intelligence algorithms need large amounts of voice data for training to retrieve good results. The provision of large amounts of operational speech data as training material is therefore essential as an enabler for an effective collaboration. However, valid and feasible operationalizations of data privacy regulations that enable the sharing of such data do not exist at the moment. As voice is a personal characteristic, it is per se biometric, i.e. it could be used to identify an operator. Therefore, the handling of such sensitive data needs to be properly organized and managed. Involved people need to be accurately informed. A dialogue needs to be initiated to discuss and clarify any worries such as any unwanted performance supervision or other misuse. In the same way, the awareness for the potential merits of the technology for safety and productivity needs to be sharpened within the company and beyond to achieve a reinforcement of the overall R&D effort for the application of speech recognition technology in ATC.

Acknowledgments

This article reports about data from a project, which has received funding from the SESAR Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement No 734141.

Disclaimer

This article reflects only the authors' view. The SESAR JU is not responsible for any use that may be made of the information it contains.

Abbreviations

AFS	Advanced Function Simulator
ANSP	Air Navigation Service Providers
ASR	Automatic Speech Recognition
ATC	Air Traffic Control
ATCO	Air Traffic Controller
ATM	Air Traffic Management
ATS	Air Traffic Services
CARS	Controller Acceptance Rating Scale
CV, CCV	Controller Clearance Verification
CWP	Controller Working Position
DFS	Deutsche Flugsicherung GmbH
EATMA	European Air Traffic Management Architecture
FDPS	Flight Data Processing System
HMI	Human Machine Interface
ICAO	International Civil Aviation Organization
ISA	Instantaneous Self-Assessment
OOG	Out Of Grammar
PSS	Paperless Strip System
PTT	Push To Talk
R&D	Research and Development
SASHA	Situation Awareness for SHAPE
SESAR	Single European Sky ATM Research program
T-FDPS	Tower Flight Data Processing System
TLA	Target Location Assistance
VCS	Voice Communication System
VOIP	Voice Over IP
VRR	Voice Recognition and Response

References

- Aerojournal (2013): Towerlotsen: Spracherkennung am Arbeitsplatz. URL retrieved on January 31, 2020: <http://www.aerojournal.de/wp-content/uploads/kalins-pdf/singles/towerlotsen-spracherkennung-am-arbeitsplatz.pdf>
- Bierwagen, T. & Vielhauer, J. (2000). Integration von Spracherkennung in die Mensch-Maschine-Interaktion am Fluglotsenarbeitsplatz, *SE im Fokus*, 2/2000, pp. 7-15, Langen: Deutsche Flugsicherung.
- Bürvenich, D. (1994). Zwischenbericht über die Erprobung von Systemen zur Spracherkennung. *DFS-interner Bericht*. Offenbach: DFS.
- Buxbaum (2012). Es hört doch jeder nur, was er versteht. *DFS-internal workshop presentation*.
- Buxbaum, J., Helmke, H., Rühl, O. & Slotty, M. (2015). Spracherkennung im Flugsicherungsumfeld. *Innovation im Fokus*, 1/2015, pp. 23-29.
- DFS (2019). ATM/ATN Masterplan. Integrierte betrieblich/technische Mittelfristplanung des Ressorts Betrieb 2020-2024, Version 1. Langen: DFS Deutsche Flugsicherung GmbH.
- DFS (2020a). Manual of Operations Air Traffic Services MO-ATS 15 Nov 2012. 19th supplementary delivery. Langen: DFS Deutsche Flugsicherung GmbH.
- EN ISO 9241-11:2018, Ergonomie der Mensch-System-Interaktion - Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte (DIN ISO 9241-11:2018).
- EUROCONTROL (2010). European Operational Concept Validation Methodology. Version 2. Brussels: EUROCONTROL.
- Helmke, H., Slotty, M., Poiger, M., Herrer, D.F., Ohneiser, O., Vink, N., Cerna, P., Hartikainen, P., Josefsson, B., Langr, D., García Lasheras, R., Marin, G., Mevatne, O.G., Moos, S., Nilsson, M.N. & Boyero Pérez, M. (2018). Ontology for transcription of ATC speech commands of SESAR 2020 solution PJ.16-04, in *IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*, London, United Kingdom.
- Helmke, H. (2019). Personal communication, SESAR Innovation Days 2019, Athens.
- ICAO (2016). Air Traffic Management. Procedures for Navigation Services, Doc 4444, Ed. 16. Montréal, Canada.
- Konopka, J. (2013). Innovativer Airport (iPort): Schlussbericht der DFS; Luftfahrtforschungsprogramm 2009-2012 (LuFo IV 2. Call); Berichtszeitraum: 01.01.2009 - 31.12.2012, Version 1.1, Langen: Deutsche Flugsicherung GmbH.
- Psychtoolbox-3 (2020). Psychophysics Toolbox Version 3 (PTB-3). URL retrieved on January 27, 2020: <http://psychtoolbox.org/>
- SESAR (2014). Innovation Analysis Report 2013. Deliverable 10.10.02-D93.
- SESAR (2019a). Controller working position/Human machine interface - CWP/HMI. URL retrieved on January 31, 2019: <https://www.sesarju.eu/projects/cwphmi>
- SESAR (2019b). SESAR 2020 PJ.16.04.02 TRL4 TVALR – Automatic Speech Recognition (ASR). Deliverable D3.2.020. Appendix A.
- Tauss, G. (2017). Die F&E-Programme SESAR 1 und SESAR 2020 – Rückblick und Ausblick. *Innovation im Fokus*, 1/17, pp. 5-8, Langen: Deutsche Flugsicherung.
- Slotty, M. & Rühl, O. (2012). Spracherkennung findet Einzug in der DFS. *TE im Fokus*, 2/2012, pp. 38-46, Langen: DFS Deutsche Flugsicherung GmbH.
- UFA (2011). DFS Enhances NEWSIM ATCoach simulator at Munich Centre with UFAs ATVoice Voice Recognition and Response. ATC network, URL retrieved on January 31, 2020: <http://www.atc-network.com/atc-news/dfs-enhances-newsim-ATCoach-simulator-at-munich-centre-with-ufas-atvoice-voice-recognition-and-response>
- UFA (2018). ATVoice. An ATC Simulation Product. URL retrieved on February 3, 2020: <http://www.ufainc.com/wp-content/uploads/2013/04/201209-UFA-ATVoice.pdf>
- UFA (2020a). homepage, URL retrieved on March 4, 2020: <https://ufainc.com/about/>
- UFA (2020b). UFA, Inc. Delivered ATVoice™ to DFS in Support of SESAR 2020 Research, press release, URL retrieved on January 31, 2020: <http://mailing.atc-network.com/t/r-l-jhziujl-diyhirdtut-h/>

ALPACA – Artificial Learning to Predict Aircraft Climb Abilities

Robin Weiß & Dr. Matthias Poppe

Abstract

To improve airspace capacity and operational efficiency, it is one key interest to predict an aircraft's horizontal and vertical flight path as precisely as possible. An Artificial Intelligence (AI) approach was used to predict an aircraft's Flight Level (FL) after six minutes based on information from only one arbitrary point in the departure phase.

A Deep Feedforward Neural Network was trained to classify such predictions into altitude groups of 2000 feet (ft) intervals, using only Mode-S and flight-plan data. Different ensemble methods were applied and tested on a pre-recorded dataset.

The best model was an ensemble model consisting of several neural networks, each specialized on a particular departure aerodrome, respectively. This model achieved an overall prediction accuracy of 69%, thus scoring 39.5% better than the BADA model, a state-of-the-art method to calculate aircraft's trajectories, on the same test dataset [1].

Introduction

To cope efficiently with the increasing traffic counts in European airspace, there is a rising need for controller assistance tools to raise capacity and safety limits. The reason is that naturally, the first ATC reaction to situations of high traffic workload is making more defensive decisions in traffic planning. This affects especially departing flights. While aircraft in descend are able to comply with a big range of vertical speed instructions, climbing flights are subject to performance limitations due to finite engine power.

In the Erlangen sector at Upper Area Control Karlsruhe, for example, Munich departures towards Northern Europe enter the sector in the vicinity of Nuremberg passing roughly FL 300 in the climb. Usually, they meet the next sector boundary to Saale sector after five to six minutes at approximately FL 360 depending on their final requested FL. Unfortunately, FL 365 is also the division level of Saale sector if it is divided into Saale low and Saale top in phases of high traffic. The traffic situation is shown in Figure 11. Additionally, these departures often meet crossing traffic travelling from Czech airspace towards France or Spain whilst restricting the Munich departures in their continuous climb. In these situations

and under high workload, departing aircraft are most likely to be planned for Saale low sector, even if they could also reach a higher FL at the Erlangen-Saale boundary. This decision imposes additional workload to both the Saale low sector as well as to the flight crew which has to stop its climb at an intermediate FL, thereby exploiting sector capacities of Saale low while leaving Saale top capacities unused.

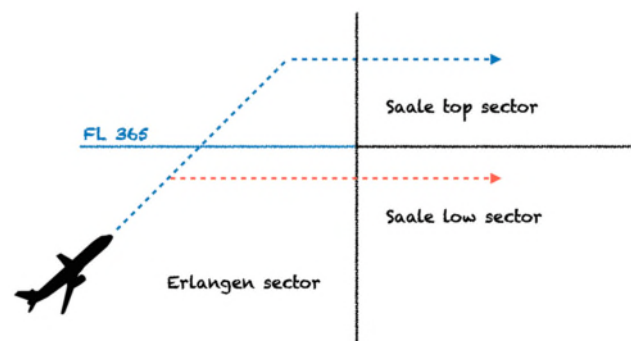


Figure 11: Munich departure climbing in Karlsruhe's Erlangen sector towards Saale sector in the North. Dashed blue line: Optimum unrestricted climb trajectory. Dashed orange line: Likely flight path in high-workload situations without quickly accessible knowledge about the aircraft's predicted climb performance, defined by the FL which can most likely be reached after six minutes.

It is very likely that similar situations are common to happen in German airspace many times a day. Therefore, it is a key interest of the DFS to develop tools to most accurately predict aircraft climb behavior in a time frame of about six minutes (parameter to be adjusted) to assist controllers in efficiently planning departure traffic on the tactical level.

We use Deep Learning with the help of Artificial Neural Networks to predict the FL which an aircraft is able to reach after six minutes, which is a typical sector transit time in German Lower airspace. The work follows up on the results concerning vertical trajectory prediction with Machine Learning (ML) published in *Innovation im Fokus* 01/18 [2] and 01/19 [3]. Note that while Deep Neural Networks (DNNs) are already being used in strategic traffic planning [4], this project focuses on a short-term, tactical application.

Dataset and Features

The dataset consists of more than 1.9 million samples which were extracted out of recorded radar data captured in 2018. Each sample describes one state of a climbing aircraft during its departure phase by the transmitted Mode-S data. From each of the more than 100,000 recorded flights, 16 samples were drawn randomly distributed over the departure phase. For a detailed description of data acquisition, extraction and filtering see [2].

The features provided for DNN training are shown in Table 2 below and are a slight modification of the feature set used in [3], table 1. The time elapsed since take-off was added to provide the DNN information with the hitherto climb performance in combination with the current FL of the aircraft. A categorical “range” feature expanded the feature set to define short, medium and long-range flights depending on the location indicators of the departure-destination pair. Two features containing the gradients of the aircraft’s current rate of climb (ROC) and indicated airspeed (IAS) were omitted to simplify the feature set.

Table 2: Features for DNN training.

Feature	Unit
Current ROC	[ft/min]
ROC peak at take-off	[ft/min]
Current FL	-
Time elapsed since take-off	[s]
Current IAS	[kn]
IAS immediately after take-off (V_2)	[kn]
Off-level flag	boolean
Range	categorical
Aircraft type	categorical

Trial runs with each time ignoring one of the features were conducted on the standard DNN. By comparing the relative drop in DNN accuracy per omitted feature, the contribution of each feature could be determined individually. Figure 12 shows the results of comparing the relative importance of the individual features.

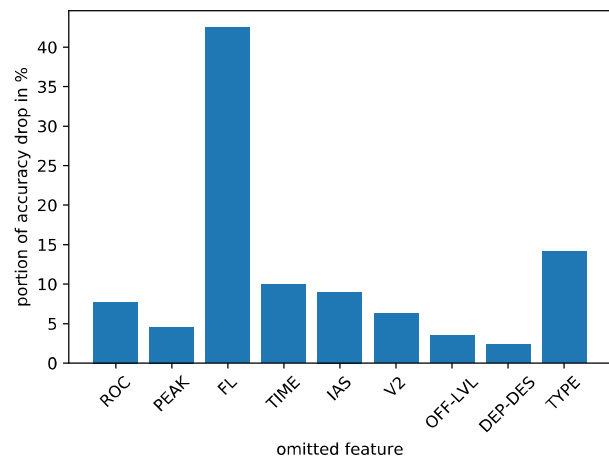


Figure 12: Contribution of the individual features measured by the relative drop of DNN accuracy when omitting this feature.

Not surprisingly, the most important information for the DNN is the current FL of the aircraft. The departure-destination pair (“DEP-DES” in Figure 12) does not provide a decisive benefit for the network. It can be assumed that the number of possible combinations between aircraft type and range is too high for the DNN to identify robust correlations about which range is to be considered as long or short for which aircraft type.

Machine Learning Methods

A basic Feedforward Deep Neural Network was set up to realize the FL prediction as a classification task. Artificial Neural Networks are mathematical models which are inspired by the structure of information processing in the human brain. As information is passed on through a number of different layers from the input stage to the output layer, the method is called “deep” learning. The term “feedforward” states that there are only single-direction connections between the nodes in the network which are pointing from the input layer towards the output layer. Our model is a compact network with only two hidden layers and a maximum width of 40 units per layer. For more information on DNNs, see [5], chapter 6. Roughly speaking, “classification” means that the prediction output is not an absolute FL number, but each example is assigned to its corresponding target FL interval, the so-called “class”, as explained in Figure 13.

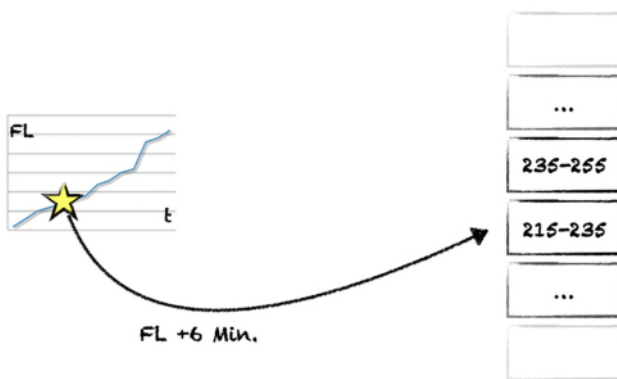


Figure 13: Core concept of the FL-prediction classification task: at one random point during the departure phase of an aircraft, the FL after six minutes is predicted as a target class defined by a certain FL band.

As we are working on recorded data, the target distribution of the labels is already known and shown in Figure 14. Each sample is annotated with an according label and the DNN learning is thus called “supervised”. Because the majority of the examples have target FLs between FL 195 and FL 295, these FLs were chosen as lower and upper bounds. Note that those values are a result of the distribution of the recorded data, while the example given in the introduction above was for explanatory purposes only.

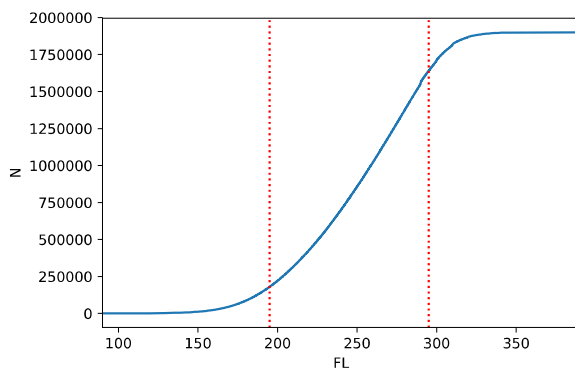


Figure 14: Distribution of target FLs (or sample labels). The red dashed lines are the bounds for the class allocation.

To limit the number of target classes to a reasonable value, five equally sized intervals were created between the lower and upper boundary. These intervals had an absolute width of 2000 ft and the final target classes for DNN training and testing were:

- [Class 1] FL 195 – FL 215
- [Class 2] FL 215 – FL 235
- [Class 3] FL 235 – FL 255
- [Class 4] FL 255 – FL 275
- [Class 5] FL 275 – FL 295

The choice of a broad class width of 2000 ft is reasonable for the benefit of a lower number of classes, because our approach assumes that it is operationally more important to obtain a reliable prediction rather than a precise FL value. The classification method offers a variety of possibilities to examine the DNN prediction’s confidence level.

Ensemble Methods

One key enhancement of the chosen approach is to make predictions by using an ensemble of models instead of only one. These individual models are specialized on frequent example characteristics which might have a systematic impact on climb performance. Aircraft type and departure aerodrome were chosen as probable candidates for such characteristics. For example, in the departure-aerodrome ensemble method, there is one model responsible to predict Frankfurt departures only, another model is specialized on flights departing from Munich, and so on. All other examples of which there are less than a certain number of instances were grouped together as an additional, mixed group with their own model. The minimum number of examples for each departure aerodrome for obtaining its own group was set to 5000, as proposed in the literature ([5], section 1.2.2). Since the ensemble method specialized on individual aircraft types did not score better than a standard model with the categorical aircraft type feature, only the departure-aerodrome ensemble method will be further discussed.

A second ensemble method follows the so-called “One-vs.-All” approach: The number of models in the ensemble is now equal to the number of target classes. The models are trained on all examples and specialized on one of the target classes. Hence, each model is responsible for assessing if a particular sample belongs to its very own class or not. In our case, this means that instead of a five-class multiclass classification, five times a binary classification (“yes or no”) is carried out. The output of each model, however, is a confidence value between 0 and 100 %, indicating that a particular sample belongs to its class. The prediction result is the class for which the corresponding model yielded the highest confidence value.

Results

The classification results are given in prediction accuracy per class. The percentage value matches the fraction of correctly classified examples in the test set. For model assessment, 75 % of the data were used for training the models, while 25 % were held back for testing. This means all tests were conducted on new,

unseen examples. The routine was carried out ten times for each method to avoid random biases in the test set. The results of the ten trials were checked for low Standard Errors of less than ± 1 %. Consequently, the low risk of having e.g. only the five most common aircraft types in the test set by chance was eliminated. It was also verified that the procedure of taking 16 examples from one flight's trajectory and making it thereby possible that examples from the same flight could occur in the training as well as in the test dataset did not induce any illicit advantage for the algorithm.

The prediction results of only the standard model without using ensemble techniques are shown in Figure 15. These results allow a first impression of the overall performance of the algorithm to enable comparing and assessing the ensemble methods afterwards. The bars show the percentage of correctly classified examples per class, the x-Axis shows FLs which separate the classes.

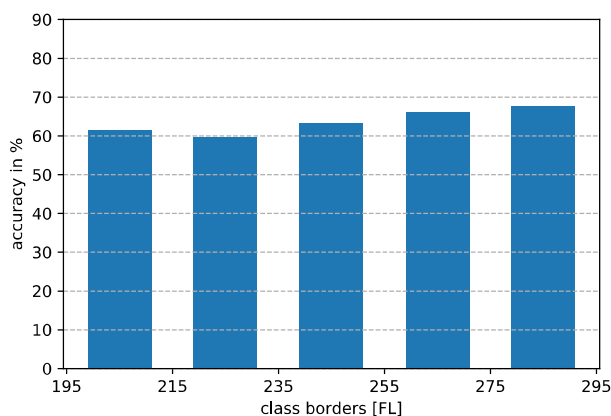


Figure 15: Prediction results of the standard model. The accuracy value is the fraction of correctly classified examples in the respective target class. The reason for the better accuracy values in higher FL classes is the comparatively higher number of training examples available in these classes.

When using the aerodrome-ensemble method with an individually specialized model for each aerodrome with enough corresponding examples, the accuracies of all classes rise to at least 60 %. The results are depicted in Figure 16 (dark blue bars). The light blue bars show the accuracy change if the prediction time parameter is reduced to three minutes only. This comparison of DNN behavior is important, because the time parameter might be subject to further adjustments. The shape of the bar tops is inverted in this case and in descending order now, because there are more training examples in the lower FL target classes after only three minutes.

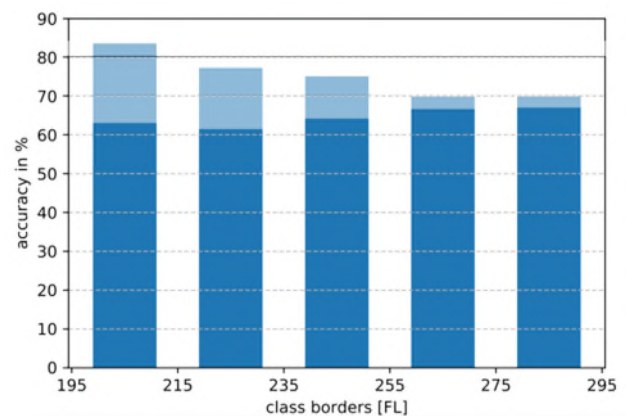


Figure 16: Prediction results of the departure-aerodrome ensemble method. In all target classes, the algorithm scores more than 60 % accuracy. Dark blue bars: Standard time parameter value of six minutes prediction horizon. Light blue bars: Accuracy results with time parameter reduced to three minutes only.

To evaluate the achieved results, Figure 17 shows the prediction accuracy of the departure-aerodrome ensemble method with six minutes prediction horizon in comparison to two other state-of-the-art trajectory prediction methods. The orange bars are the results of EUROCONTROL's BADA 3 database, which is able to calculate an aircraft's whole trajectory based on a point-mass model of the aircraft including aerodynamics as well as atmospheric models [6]. However, BADA uses a rigid set of formulae which does not include AI or ML. Note that the presentation of the BADA results in target classes as in Figure 17 is for comparison purposes only. As a standalone model, the BADA prediction output is not a class number but an absolute FL value and furthermore the aircraft's trajectory to that point and beyond.

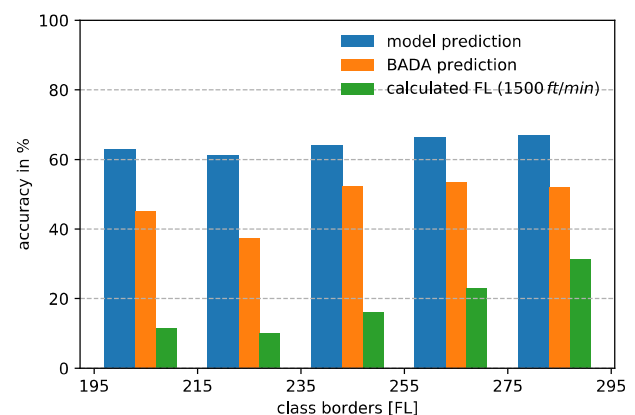


Figure 17: Comparison of the departure-aerodrome ensemble method result with two different state-of-the-art prediction methods.

The green bars in Figure 17 show the results of the simple calculation of a static mean rate of climb of 1500 ft/min as it has been done in previous ATM systems. The result of the calculation is, like in the BADA case, from the absolute FL value projected into the target classes for comparison. Taking the mean accuracy over all classes, the DNN model scores 39.5 % better than the BADA prediction and more than 200 % better than the calculation with 1500 ft/min.

Another interesting behavior occurs when the number of classes is set to ten instead of five target classes, resulting in a class width of 1000 ft, equal to the vertical separation minimum between aircraft. While the overall mean prediction accuracy of the DNN drops significantly from about 64 % to roughly 43 % as expected, the DNN is now even 62.9 % better than the respective projection of BADA predictions into ten classes.

Lastly, a different method is the One-vs.-All approach. Now, the number of models is equal to the number of target classes. Each model is a DNN which is trained on all examples in the training dataset, but only making a binary decision whether a particular sample belongs to its own class or not. In this case, the final prediction is the class which exhibits the highest prediction confidence value. The output of the whole model, though, is a vector with confidence levels for the sample belonging to the respective target class. Having not only the predicted class number but also the corresponding confidence level as prediction output can be of pivotal importance for using such systems in a high-reliability environment such as ATC.

With about 64 % correct predictions, the overall balanced accuracy of the One-vs.-All method is almost equal to the performance of the departure-aerodrome ensemble method. Figure 18 characterizes the behavior of the prediction accuracy when demanding a specific confidence level. For example, only correct predictions with a classifier confidence of more than 90 % are counted as correct predictions, while correct predictions with a lower confidence level are refused. Note that for simplicity purposes, the overall accuracy was used instead of the balanced accuracy [7]. Hence with 69.1 %, the baseline value is slightly higher than the balanced accuracy given for comparison above. Even if the demanded confidence level is a high value such as 90 %, the calculated classifier performance is 48 %, which means that almost every second aircraft's FL class is still predicted correctly and with the respective confidence level.

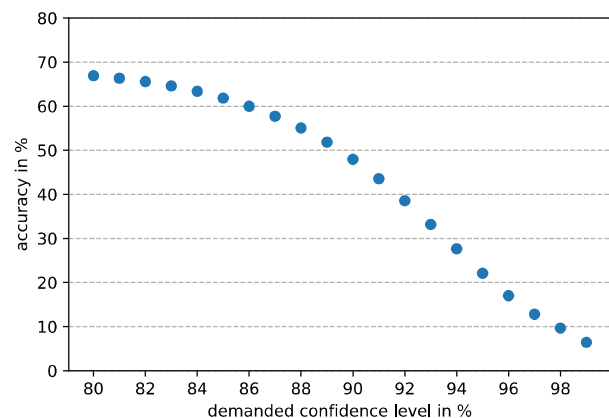


Figure 18: One-vs.-All classifier accuracy level in the demanded confidence level interval [80, 99] %. The y-axis results are the fraction of correctly classified examples with a prediction confidence equal or higher to the x-axis value at the same time.

It is also crucial to look at interpretations of different possible prediction errors. An overestimation of the FL prediction with high confidence will most likely have a more negative effect than a less-confident underestimation. Relying on an allegedly confident prediction to plan an aircraft at a too high FL after six minutes results in planning a trajectory which is operationally not realizable for the aircraft. Otherwise, a too low prediction will limit efficiency and potentially result in a wrong sector sequence (cf. Figure 11), but at least not in planning an impossible trajectory. When admitting too low predictions for example in the One-vs.-All method, the respective overall accuracy rises from about 69 % to roughly 85 %, which is a remarkable prediction quality.

Discussion

The results show that with a compact DNN and instantaneously obtained Mode-S data, it is possible to set up a stand-alone framework to provide the basic requirements for a quick and easy ATC assistance tool for the prediction and planning of vertical aircraft movements. However, it must be kept in mind that our approach does not cover the creation of aircraft's trajectories neither whole nor in parts, as it can be done with BADA. There are other approaches as well, which enhance the BADA results by using AI and ML [8], but do not provide stand-alone methods. Other research detaches from BADA but applies its methods only on common, frequent aircraft types [9]. Our analysis showed that the existence of uncommon aircraft types in our more comprehensive dataset poses an upper limit to prediction accuracy. At the same time, experiments suggest that this limit can yet be overcome by simply

collecting more training examples from these aircraft types.

To further improve prediction quality of our own methods, the feature set should be expanded to some aspects which have not been considered yet:

- Meteorological data, particularly the wind gradient influences the aircraft's climb performance.
- The airline operator code might provide conclusions about the operator's Cost Index strategy.
- Information about ATC restrictions or clearances interfering with the aircraft's usual unrestricted climb behavior are only marginally included at the moment by the "Off-level flag" feature but should be implemented in greater detail.
- Knowledge about which flight route is to be considered short, medium or long range for a particular aircraft type could give an important clue about the aircraft's expected mass.

Additionally, the assessment of prediction quality and, in particular, reliability of such methods is an issue which requires critical consideration. It would be desirable to create AI methods which are able to evaluate before the actual prediction if, for the given flight, a reliable prediction is possible, or methods which are more robust against unexpected data or "outliers". This might for example be achieved by training a DNN explicitly to prefer more defensive predictions.

For the near future, it might be worth to continue training the DNN with live radar data. This way, e.g. the sample size for less common aircraft types could be increased and further analysis of prediction behavior and quality could be conducted to gain better knowledge about ML applications in ATC operations.

Abbreviations

AI	Artificial Intelligence
ATC	Air Traffic Control
ATM	Air Traffic Management
BADA	Base of Aircraft Data
DEP-DES	Departure-Destination
DFS	Deutsche Flugsicherung GmbH
DNN	Deep Neural Network
FL	Flight Level

IAS	Indicated Airspeed
ML	Machine Learning
ROC	Rate of Climb

References

- [1] Weiß, R. (2020). Investigating Classification Methods to Predict Aircraft Climb Behaviour. Master's Thesis, University of Osnabrück.
- [2] Fieberg, D. & Poppe, M. (2018). Machine Learning Feature Extraction for Improved Trajectory Prediction. Innovation im Fokus, 01/18.
- [3] Poppe, M., Pütz, T. & Scharff, R. (2019). Improving Climb Performance Prediction in Air Traffic Control with Machine Learning and Full Flight Simulator Verification. Innovation im Fokus, 01/19.
- [4] EUROCONTROL (2018). Predicting Flight Routes with a Deep Neural Network in the Operational Air Traffic Flow and Capacity Management System. Retrieved online, April 2020: <https://www.eurocontrol.int/sites/default/files/2019-06/traffic-prediction-improvements-tpi-factsheet.pdf>
- [5] Goodfellow, I., Bengio, Y. & Courville, A. (2016). Deep Learning. Cambridge: MIT Press.
- [6] Poles, D., Nuic, A. & Mouillet, V. (2010). Advanced Aircraft Performance Modeling for ATM: Analysis of BADA Model Capabilities. Retrieved online, April 2020: https://www.eurocontrol.int/eec/gallery/content/public/document/eec/conference/paper/2010/006_Analysis_of_BADA_modelling_capabilities.pdf
- [7] Sokolova, M. & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing and Management, 45(4), 427-437.
- [8] Alligier, R. & Gianazza, D. (2018). Learning Aircraft Operational Factors to Improve Aircraft Climb Prediction: A Large Scale Multi-Airport Study. Transportation Research Part C: Emerging Technologies, 96, 72-95.
- [9] Alligier, R. & Gianazza, D. (2015). Machine Learning and Mass Estimation Methods for Ground-Based Aircraft Climb Prediction. IEEE Transactions on Intelligent Transportation Systems, 16(6), 3138-3149.

Innovation im Fokus Informationen zu Forschung, Entwicklung und Validierung

Innovation im Fokus erscheint halbjährlich und beschäftigt sich bevorzugt mit Informationen zu Forschung, Entwicklung und Validierung mit Beteiligung DFS Deutsche Flugsicherung GmbH. Diese Ausgabe ist elektronisch im Internet (www.dfs.de > [Flugsicherung](#) > [F&E](#) > [Service](#)) sowie über das DFS Intranet verfügbar. 90 Exemplare werden in gedruckter Form aufgelegt.

Wie auch der Vorgänger-Zeitschrift „TE im Fokus“ wurde dieser Zeitschrift von der Deutschen Bibliothek eine ISSN (International Standard Serial Number) zugeteilt:

Printversion: ISSN 2198-8951 (vormals: 1861-6364)

Internet-Version: ISSN 2198-896X (vormals: 1861-6372)

Datum dieser Ausgabe: 29.06.2020

DISCLAIMER

Alle hier erwähnten Warenzeichen sind Eigentum ihrer jeweiligen Besitzer. Warenzeichen werden nicht ausdrücklich als solche gekennzeichnet. Aus dem Fehlen von Urheber- oder Markenrechtskennzeichen darf jedoch nicht geschlossen werden, dass es sich um einen nicht geschützten Namen oder um eine nicht geschützte Marke handelt.

COPYRIGHT

Dieses Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwendung außerhalb der Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung der DFS Deutsche Flugsicherung GmbH unzulässig und wird zivil- und strafrechtlich verfolgt. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Copyright © 2020 by DFS Deutsche Flugsicherung GmbH - Weitergabe sowie Vervielfältigung dieser Unterlage, Verwertung und Mitteilung ihres Inhalts nicht gestattet, soweit nicht ausdrücklich zugestanden. Zuwiderhandlungen verpflichten zu Schadensersatz. Alle Rechte vorbehalten, insbesondere für den Fall der Patenterteilung oder GM-Eintrag.



Impressum

Innovation im Fokus

Informationen zu Forschung,
Entwicklung und Validierung

Herausgeber:

DFS Deutsche Flugsicherung GmbH

Günter Achatz, Bereichsleitung
Planung & Innovation

Redaktion:

Dr. Konrad Hagemann
Tel. +49 (0)6103 707 5745
E-Mail: konrad.hagemann@dfs.de

Stefan Tenoort
Tel. +49 (0)6103 707 5769
E-Mail: stefan.tenoort@dfs.de

Dr. Morten Grandt
Tel. +49 (0)6103 707 1139
E-Mail: morten.grandt@dfs.de

Oliver Haßa
Tel. +49 (0)6103 707 5762
E-Mail: oliver.hassa@dfs.de

Iris Filbrich
Tel. +49 (0)6103 707 1131
E-Mail: iris.filbrich@dfs.de

Anschrift der Redaktion:

DFS Deutsche Flugsicherung GmbH
Redaktion Innovation im Fokus
Am DFS-Campus 5
63225 Langen
E-Mail: forschung@dfs.de

Nachdruck nur mit Genehmigung.

