



Erkennung gesprochener Flughafen-Codes

Studienarbeit am Institut für Logik, Komplexität und Deduktionssysteme

Prof. Dr. A. Waibel

Fakultät für Informatik

Universität Karlsruhe (TH)

von

cand. inform.

Michael Dambier

Betreuer:

Prof. Dr. A. Waibel

Dr. I. Rogina

Tag der Anmeldung: 01. April 2002

Tag der Abgabe: 28. Juni 2002

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Karlsruhe, den 28. Juni 2002

Inhaltsverzeichnis

1	Einleitung	1
1.1	Zielsetzung der Arbeit	2
1.2	Gliederung der Arbeit	2
2	Flugtechnische Grundlagen	3
2.1	Flugnavigation	3
2.2	Aufbau der Flughafen - Codes	4
2.3	Datenbank der Flughafen - Codes	7
3	Der Spracherkenner	11
3.1	Spracherkenner mit N-Gramm-Modell (Buchstaben)	11
3.2	Spracherkenner mit N-Gramm-Modell (vollständige Codes)	12
3.3	Spracherkenner mit CFG-Modell	13
4	Evaluierungsergebnisse	17
4.1	Evaluierung mit geräuscharmen Sprachdaten	17
4.1.1	Spracherkenner mit N-Gramm-Modell (Buchstaben)	18
4.1.2	Spracherkenner mit N-Gramm-Modell (vollständige Codes)	19
4.1.3	Spracherkenner mit CFG-Modell	20
4.2	Evaluierung mit geräuschbehafteten Sprachdaten	22
4.2.1	Spracherkenner mit N-Gramm-Modell (Buchstaben)	22
4.2.2	Spracherkenner mit N-Gramm-Modell (vollständige Codes)	22
4.2.3	Spracherkenner mit CFG-Modell	23
5	Zusammenfassung und Ausblick	25
A	Aufbau und Benutzung der Demonstrationsversionen	27
	Literatur	33

1. Einleitung

Sprachgesteuerte Flugnavigationssysteme für Flugzeuge und Helikopter gibt es bereits in der militärischen Luftfahrt. Diese Systeme sind leider nicht sprecherunabhängig, sondern müssen vorher aufwändig (über mehrere Stunden) von den Piloten trainiert werden. In militärischen Anwendungen ist dies möglich, da meist nur ein oder zwei Piloten ein Flugzeug oder Hubschrauber über einen längeren Zeitraum fliegen. Des weiteren müssen mehrere Kommandos, auf die das Navigationssystem reagiert, von den Piloten erlernt werden.

In der Allgemeinen Luftfahrt werden sprachgesteuerte Navigationssysteme noch nicht eingesetzt, da zum Teil mehrere Piloten ein Flugzeug fliegen und jeder Pilot somit Stunden vor dem Abflug das Spracherkennungssystem trainieren und sich mit den Befehlen, die das System versteht, auseinandersetzen müsste. Dies ist nicht wirtschaftlich und in diesem Sinne auch nicht praktikabel. Es wäre also wünschenswert, wenn Piloten ohne vorheriges Training mit dem System sprechen bzw. navigieren könnten.

Dies wäre möglich, wenn der Pilot das System einmal trainieren und das Ergebnis dieses Trainingsprozesses dem Piloten auf einer Diskette, Chipkarte oder einem anderen Medium ausgehändigt werden würde. Dieses Vorgehen erfordert allerdings einen herstellerunabhängigen, einheitlichen Standard der Datenstrukturen, in denen die akustischen Daten des Piloten gespeichert werden und mit denen zahlreiche kommerzielle Realisierung arbeiten können. Anderenfalls, bei uneinheitlichen Standards, bleibt es nicht auf ein Training und das Mitführen eines Mediums beschränkt. Das Medium bzw. die mitzuführenden Medien sollten den Anforderungen des Pilotenalltags standhalten und der Pilot dürfte es bzw. sie nicht zu Hause vergessen. Dies erscheint mir, den Arbeitsalltag eines Piloten etwas näher betrachtet, wenig praktikabel.

Es stellt sich die Frage, ob nicht ein handelsüblicher Diktiererkenner dies alles leisten könnte. Der Trainingsaspekt wurde bereits im vorhergehenden Absatz angesprochen. Dieser wäre weiterhin bei solchen Erkennern vorhanden. Ein Spracherkenner, dessen Einsatzbereich das Cockpit ist, muss in stark geräuschbehafteten Umgebungen beim Betrieb eines Luftfahrzeuges genauso gut funktionieren wie in ruhiger, geräuschloser

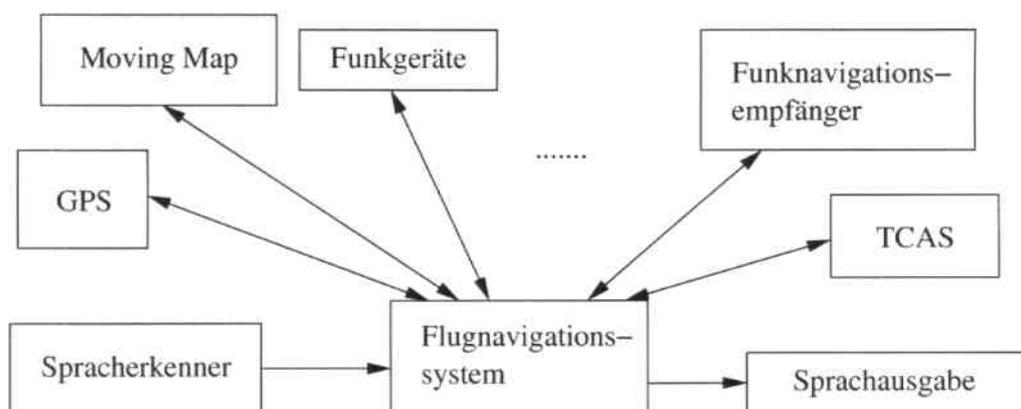


Abbildung 1.1: Interaktion des Flugnavigationssystems mit anderen Komponenten

Umgebung. Dies vermag bis jetzt kein handelsüblicher Spracherkenner zu leisten. Hierzu sind spezielle Ansätze bzw. Algorithmen erforderlich.

1.1 Zielsetzung der Arbeit

Diese Arbeit befasst sich mit dem Versuch, einen trainierten, sprecherunabhängigen Spracherkenner als Grundlage für ein Flugnavigationssystem einzusetzen. Hierzu sollten Spracherkenner mit unterschiedlichen Sprachmodellen, die für die Flugnavigation wichtige, buchstabierte Flughafen-Codes erkennen, aufgebaut und getestet werden. Ziel ist es, ein geeignetes Spracherkennungssystem zu finden, welches mit weiteren Modifikationen als späteres Dialogsystem für die Flugnavigation in einem Helikopter eingesetzt werden kann. Dieses soll den Piloten administrative und flugnavigationstechnische Arbeit durch Kommunikation mit flugtechnischen Geräten abnehmen. Dabei soll die verwendete Sprache Englisch sein, da sich auf der ganzen Welt Englisch als einheitliche Sprache in der Luftfahrt etabliert hat. Ein Überblick über die unterschiedlichen Komponenten¹ des Flugnavigationssystems ist in Abb. 1.1 dargestellt.

1.2 Gliederung der Arbeit

Im folgenden Kapitel 2 werden zuerst die für das Verständnis wichtigen luftfahrt-spezifischen Grundlagen erläutert. Diese sind die Flugnavigation, der Aufbau der Flughafen-Codes, das internationale Buchstabieralphabet und der Aufbau der Datenbank für Flugplatzinformationen. In Kapitel 3 wird das initiale, zugrundeliegende Spracherkennungssystem und der Aufbau der verschiedenen Spracherkenner aus dem Initialsystem erläutert. Danach werden in Kapitel 4 die Evaluierungsergebnisse der Spracherkenner diskutiert. Eine Zusammenfassung und ein Ausblick auf die Zukunft der Technologie in der Allgemeinen Luftfahrt in Kapitel 5 beenden die Ausführungen. Im Anhang A wird eine Einführung in den Aufbau und die Benutzung der Demonstrationsversionen der Spracherkenner gegeben.

¹TCAS – Traffic Alert and Collision Avoidance System

2. Flugtechnische Grundlagen

In diesem Kapitel wird eine Einführung in die Luftfahrt gegeben. Diese umfasst die Flugnavigation sowie den Aufbau der Flughafen-Codes und den Aufbau der Datenbank mit den Flugplatzinformationen.

Beim Fliegen eines Flugzeuges müssen das Steuerhorn bzw. der Steuerknüppel, je nach Bauart des Flugzeuges, für das Höhen- und Querruder mit einer Hand, die Fußpedale für das Seitenruder und ab und zu der Gashebel mit der anderen Hand bedient werden. Somit ist meistens eine Hand frei, um notwendige funknavigatorische Einstellungen vornehmen zu können. Beim Hubschrauber ist das etwas anders. Dort bewegt eine Hand den Steuerknüppel für den Hauptrotor zur Bestimmung der Flugrichtung, die andere Hand den Blattverstellhebel für vertikale Bewegungen und die Füße die Pedale zur Steuerung des Heckrotors. Somit ist während des Fluges keine Hand mehr frei, mit der man Einstellungen vornehmen kann. Zur Zeit nimmt entweder der Copilot die Einstellungen vor, oder der Pilot läßt entweder Steuerknüppel oder Blattverstellhebel los, um dies selbst zu tun. Dieses Loslassen stellt aber ein Risiko der Flugsicherheit dar und sollte nicht praktiziert werden. Leider bleibt dem Piloten bei Alleinflügen keine andere Wahl. Deshalb soll hier ein spracherkennendes Flugnavigationssystem zum Einsatz kommen, mit dessen Hilfe der Pilot Einstellungen bestimmter Geräte vornehmen kann und das darüber hinaus weitere Informationen für den Piloten bereithält, so dass dieser nicht von seiner Haupttätigkeit, dem Fliegen, abgelenkt wird — getreu dem Motto: „first fly the aircraft“.

2.1 Flugnavigation

Bei der Flugnavigation ist grundsätzlich nach den Flugregeln zu unterscheiden. Bei einem Flug nach Instrumentenflugregeln (IFR¹) ist der Flugweg von vorneherein festgelegt und kann durch die Angabe aller Flugnavigationspunkte beschrieben werden. Als solche Navigationspunkte dienen grösstenteils VOR², DME³ und NDB⁴.

¹IFR – Instrumental Flight Rules

²VOR – VHF Omnidirectional Range

³DME – Distance Measuring Equipment

⁴NDB – Non Directional Beacon

Das VOR ist eine Funknavigationseinrichtung die in alle Richtungen eine auf den Standort des VOR bezogene Richtungsinformation (Gradzahl bezogen auf den magnetischen Nordpol) aussendet. Damit kann man leicht eine beliebige, vorher festgelegte, Richtung bezüglich des VOR fliegen. Ein DME ist eine Einrichtung, die empfangene Funksignale zeitverzögert wieder zurücksendet. Daraus kann das Gerät im Flugzeug die Distanz (Luftlinie) berechnen. Es ist meist mit einem VOR gekoppelt, um Richtungsinformationen zusammen mit den Distanzinformationen auf einer Funkfrequenz nutzen zu können.

Gegenüber dem VOR sendet ein NDB, wie der Name „non directional beacon“ bereits aussagt, ein ungerichtetes Signal aus. Dieses kann man für direkte Anflüge auf die Navigationseinrichtung und natürlich auch für Abflüge von dieser benutzen. Allerdings ist dies aufgrund der fehlenden Richtungsinformation schwerer. Des weiteren erfolgen bei einem IFR-Flug mehrere Wechsel der Funkfrequenzen für die Flugsicherung.

Ein Pilot, der nach Sichtflugregeln (VFR⁵) fliegt, wird nur durch die Luftraumstruktur in seinem Flugweg beschränkt. Er kann VOR's, DME's oder NDB's zur Navigation nutzen, es ist aber nicht zwingend vorgeschrieben.

Bei Flügen nach beiden Flugregeln (VFR, IFR) wird mit zunehmendem Maße die GPS-Technologie eingesetzt, welche die Flugnavigation erheblich vereinfacht. Die DFS Deutsche Flugsicherung GmbH führte mit dem Projekt „EAM04“ Flächennavigationspunkte ein. Bei der Flächennavigation (RNAV⁶) werden beliebige Punkte, welche durch die Angabe der GPS-Koordinaten spezifiziert sind, für die Navigation genutzt. Diese Nutzung ist ohne ein GPS-Gerät im Cockpit nicht möglich. Da diese Verfahren verbindlich für alle Flüge nach Instrumentenflugregeln sind, ist davon auszugehen, dass immer mehr Luftfahrzeuge der Allgemeinen Luftfahrt mit GPS-Geräten ausgerüstet sind. Leider ist die Bedienung dieser Geräte so komplex, dass man sich eingehend damit beschäftigen muss, bevor man sie im Cockpit einsetzen kann. Denn durch die komplexe Bedienweise können sie den Piloten von seinen eigentlichen Aufgaben ablenken. Dies wäre ein zusätzlicher Grund für eine Sprachsteuerung des Navigationssystems. Ein sprachgesteuertes Dialogsystem versperrt nämlich nicht die Sicht auf wichtige Instrumente und den Luftraum.

2.2 Aufbau der Flughafen - Codes

Das spracherkennende Flugnavigationssystem muss unabhängig von den Flugregeln den Zielflugplatz erkennen können. Da die verwendeten Flugplatznamen öfter vorkommen können, werden Ortskennungen verwendet, die hier als Flughafen-Codes bezeichnet werden. Diese Codes sind weltweit einheitlich aufgebaut (Empfehlung der ICAO⁷) und werden heute in allen computergestützten Flugplanungssystemen und GPS-Empfängern verwendet.

Ortskennungen sind Abkürzungen, die aus vier Buchstaben bestehen. Damit werden Städte, Fernmeldeeinrichtungen, Flughäfen oder andere Einrichtungen, die für

⁵VFR – Visual Flight Rules

⁶RNAV – Area Navigation

⁷ICAO – International Civil Aviation Organisation

die Luftfahrt wichtig sind, gekennzeichnet. Sie wurden eingeführt, um Luftfahrtmeldungen, zum Beispiel Flugpläne, zu verkürzen und Verwechslungen auszuschließen.

Der Aufbau dieser Flughafen-Codes soll nun an einem Beispiel erläutert werden. Der Flughafen in Mannheim-Neustadt wird mit dem Code EDFM bezeichnet.

Der erste Buchstabe im Flughafen-Code, im Beispiel E, bezeichnet das Flugfernmeldegebiet (AFSRA⁸). Die Erde ist in disjunkte Flugfernmeldegebiete mit eigenen Identifizierungsbuchstaben unterteilt. Deutschland liegt im Flugfernmeldegebiet mit dem Identifizierungsbuchstaben E. Es fangen also alle Codes deutscher Flughäfen mit E an.

Der zweite Buchstabe, im Beispiel D, bezeichnet eigentlich den Staat, in dem der Flughafen liegt. Der Staat kann in bestimmten Grenzen den zweiten Buchstaben der Ortskennungen beeinflussen. So wurden in der Bundesrepublik Deutschland die Ortskennungen im Januar 1995 (AIC⁹ 13/94, gültig ab 5. Januar 1995) reorganisiert. Für militärische Flugplätze wurde hier der Buchstabe T festgelegt, für zivile Flughäfen, Landeplätze und sonstige Einrichtungen der Buchstabe D.

Die Bedeutung des dritten und vierten Buchstaben kann jeder Staat nach eigenem Ermessen festlegen. In der Bundesrepublik Deutschland bezeichnet der dritte Buchstabe die zuständige Flugberatungsstelle, in deren Bereich der Flugplatz oder die Einrichtung liegt. Dabei können mehrere Buchstaben einem Flugberatungsdienst zugeordnet sein.

In Tabelle 2.1 sind die jeweiligen Flugberatungsdienste mit dem zugeordneten Kennbuchstaben aufgeführt. Im Beispiel ist es der Buchstabe F — Flugberatungsdienst (AIS¹⁰) Frankfurt. Bei militärischen Einrichtungen wird in Deutschland die Regel für die Zuteilung des dritten Buchstabens zu dem jeweiligen AIS nicht angewandt. Hier kann der dritte Buchstabe ein beliebiger Buchstabe aus dem Alphabet sein.

Der vierte Buchstabe bezeichnet den Anfangsbuchstaben der Stadt, auf deren Gebiet der Flugplatz oder die Einrichtung liegt. Durch diese Zuordnung kommt es allerdings zu „Doppelbelegungen“, das heißt, es gibt einen Code mit dem zwei Flugplätze gekennzeichnet werden würden. Da ein Code aber aus Gründen der Eindeutigkeit einem Flugplatz zugeordnet werden muss, weicht man in diesen Fällen vom Anfangsbuchstaben der Stadt ab und macht Ausnahmen. Beispiele hierfür sind Düsseldorf (EDDL), Hannover (EDDV), Saarbücken (EDDR), Bremen (EDDW) und Leipzig (EDDP).

Im oben angeführten Beispiel steht EDFM also für

E Flugfernmeldegebiet E,

D ziviler Flughafen, Landeplatz oder sonstige Einrichtung,

F Flugberatungsdienst (AIS) Frankfurt und

M Stadt Mannheim.

⁸AFSRA – Aeronautical Fixed Service Routing Area

⁹AIC – Aeronautical Information Circular

¹⁰AIS – Aeronautical Information Service

Buchstabe	Flugberatungsdienst
A	AIS Berlin
B	AIS Berlin
C	AIS Berlin
D	internationaler Verkehrsflughafen
E	AIS Frankfurt
F	AIS Frankfurt
G	AIS Frankfurt
H	AIS Hamburg
I	nicht zugewiesen
J	nicht zugewiesen
K	AIS Köln
L	AIS Düsseldorf
M	AIS München
N	AIS München
O	AIS Berlin
P	AIS München
Q	AIS Nürnberg
R	AIS Saarbrücken
S	AIS Stuttgart
T	AIS Stuttgart
U	nicht zugewiesen
V	AIS Hannover
W	AIS Bremen
X	AIS Bremen
Y	nicht zugewiesen
Z	MET

Tabelle 2.1: Zuordnung des dritten Buchstaben in Ortskennungen (aus [Fran97], S.39ff)

Buchstabe	Schlüsselwort	Aussprache
A	Alpha	Alfa
B	Bravo	Brawo
C	Charlie	Tschahrli
D	Delta	Delta
E	Echo	Ecko
F	Foxtrott	Foxtrott
G	Golf	Golf
H	Hotel	Hotel
I	India	Indja
J	Juliett	Dschuljett
K	Kilo	Kilo
L	Lima	Lima
M	Mike	Maik
N	November	Nowemmba
O	Oscar	Osska
P	Papa	Papah
Q	Quebec	Kibeck
R	Romeo	Rohmio
S	Sierra	Sierra
T	Tango	Tängo
U	Uniform	Juniform
V	Victor	Wiktor
W	Whisky	Wisski
X	X-Ray	Exre
Y	Yankee	Jänki
Z	Zulu	Sulu

Tabelle 2.2: Das internationale Buchstabieralphabet (aus [Fran97], S.268)

Die Flughafen-Codes werden bei der verbalen Übermittlung nach dem internationalen Buchstabieralphabet buchstabiert, um die Übertragung einfacher und weniger fehleranfällig zu machen. Das internationale Buchstabieralphabet ist in Tabelle 2.2 mit den Buchstaben, Schlüsselwörtern und der jeweiligen Aussprache angegeben.

2.3 Datenbank der Flughafen - Codes

Um mit den Flughafen-Codes arbeiten zu können, benötigt man zuerst eine möglichst vollständige Auflistung aller deutschen Flugplätze. Unter <http://www.eddh.de> findet man unter der Rubrik „→ Equipment → Kniebrett“ eine Datei mit einer Auflistung aller deutschen Flugplätze mit Angaben wie Flughafen-Code, Name des Flughafens und der Stadt und Funkfrequenzen für die Flugverkehrskontrollstelle und die ATIS¹¹ (siehe [Wimm01]). Diese Daten wurden in eine Janus-konforme Datenbank überführt, deren Datenstruktur in Tabelle 2.3 angegeben ist.

¹¹ATIS – Automatic Terminal Information Service – Automatische Ausstrahlung von Start- und Landeinformationen auf einer Funkfrequenz durch die DFS Deutsche Flugsicherung GmbH

Name des Feldes	Erläuterung	Beispiel
icao	Flughafen-Code	EDDS
town	Name der Stadt, auf deren Gebiet der Flugplatz liegt	Stuttgart
airport	Name des Flugplatzes	STUTTGART
elevation	Höhe des Flugplatzes über dem Meeresspiegel gemessen in Fuss (ft)	1267
varDeg	Grad der Missweisung (Variation) am Flugplatz	0
varMin	Minuten der Missweisung (Variation) am Flugplatz	0
latDeg	Grad der geographischen Breite (Koordinaten) des Flugplatzes	48
latMin	Minuten der geographischen Breite (Koordinaten) des Flugplatzes	41.4
longDeg	Grad der geographischen Länge (Koordinaten) des Flugplatzes	-9
longMin	Minuten der geographischen Länge (Koordinaten) des Flugplatzes	13.3
frequency	Sprechfunkfrequenz des Flugplatzes	119.050
atis	Funkfrequenz auf der, falls vorhanden, die ATIS ausgestrahlt wird	126.120
runwayDesignation	Pistenbezeichnung (Start- und Landerichtungen)	07/25
runwayDimension	Angabe zur Länge und, falls vorhanden, zur Breite der Piste	3345x45m
runwaySurface	Belag der Piste	Concrete

Tabelle 2.3: Die Datenbankstruktur der Flugplatz-Datenbank

Die Angaben im Datenfeld „frequency“ der Datenbank müssten noch nach unterschiedlichen Sprechfunkfrequenzen auf kontrollierten Flugplätzen für Turm (Tower), Rollkontrolle (Ground), Vorfeld (Apron), etc. differenziert werden. Leider standen hierzu keine Daten zur Verfügung, so dass im Feld „frequency“ bei kontrollierten Flugplätzen¹² die Sprechfunkfrequenz für TURM (TOWER) und bei unkontrollierten Flugplätzen die Sprechfunkfrequenz für INFO angegeben ist. Aus den vorhandenen Daten ist nicht ersichtlich, ob der Flugplatz kontrolliert oder unkontrolliert ist.

Nachdem nun eine Einführung in einige Grundlagen der Luftfahrt gegeben wurde, wird im nächsten Kapitel der Aufbau der einzelnen Spracherkenner erläutert.

¹²Flugplätze mit Flugverkehrskontrolle durch die DFS Deutsche Flugsicherung GmbH und entsprechend eingerichtetem Luftraum

3. Der Spracherkenner

In diesem Kapitel soll der Aufbau der verschiedenen Spracherkenner erläutert werden. Als Basis für die Erkener wurde der Broadcast-News-Spracherkenner des ILKD¹ der Universität Karlsruhe (TH) gewählt. Dieser wurde in amerikanischem Englisch mit Texten aus Zeitungsartikeln und anderen Medien von Personen, deren Muttersprache Englisch ist, trainiert. Er basiert auf dem Janus Recognition Toolkit der Version 5.0, das an den Interactive Systems Laboratories (ISL) der Universität Karlsruhe (TH) und der Carnegie Mellon University in Pittsburgh erstellt wurde und besitzt eine Signal- und Sprecheradaption, das heisst der Spracherkenner stellt sich automatisch auf einen Sprecher ein.

Da die Spracherkenner nicht neu trainiert werden sollten, wurden die Phoneme, das Codebuch, die Distributionen, die Transition Models und die Topologie unverändert übernommen. An die neuen flugtechnischen Anforderungen wurden das Vokabular, das Wörterbuch und die verschiedenen Sprachmodelle, in denen sich die einzelnen Erkener unterscheiden, angepasst. Die einzelnen Spracherkenner werden nun in den folgenden Unterkapiteln vorgestellt.

3.1 Spracherkenner mit N-Gramm-Modell (Buchstaben)

Der erste Spracherkenner wurde auf das Erkennen von beliebigen Buchstaben, die nach dem internationalen Buchstabieralphabet buchstabiert werden, programmiert. Dies folgt der Praxis des Flugfunks. Dort werden die Flughafen-Codes immer nach dem internationalen Buchstabieralphabet buchstabiert, genauso weitere für die Flugnavigation wichtige Abkürzungen, wie zum Beispiel die Funknavigationseinrichtungen VOR, DME oder NDB. Das Wörterbuch beinhaltet dabei verschiedene Aussprachevarianten der Buchstaben, aufgebaut aus Phonemen, nach dem internationalen Buchstabieralphabet (siehe Tabelle 2.2) sowie verschiedene Umgebungsgeräusche (Müllworte). Im Wörterbuch sind alle Wörter aufgeführt, die der Spracherkenner erkennen kann, wobei jedes Wort mit einer bis fünf Aussprachevarianten enthalten ist.

¹ILKD – Institut für Logik, Komplexität und Deduktionssysteme

Zum Beispiel:

DELTA {D E H L T A X}

DELTA(2) {D E H L T A A}

Die Aussprachevarianten wurden nach einem Vergleich mit den verschiedenen Aussprachevarianten des Wörterbuchs des Broadcast-News-Spracherkenners zusammengestellt. Da das Wörterbuch nur die Buchstaben des Alphabets sowie deren verschiedene Aussprachevarianten und ein paar Störgeräusche umfasst, ist es im Vergleich zu anderen Spracherkennern sehr klein. Dies wirkt sich positiv auf die zeitliche Dauer der einzelnen Spracherkennungsläufe und auf deren Ergebnisse aus.

Im Vokabular werden wiederum die Buchstaben des internationalen Buchstabieralphabets, diesmal aber ohne Aussprachevarianten und die verschiedenen Umgebungsgeräusche aufgeführt.

Das Sprachmodell wurde auf der Grundlage aller Flughafen-Codes deutscher Flughäfen und -plätze aus der Datenbank (dies sind 445 Stück) mit Hilfe des Interactive Systems LM Toolkit (siehe [RiSG96]) erstellt. Es basiert auf Tri- bzw. Viergrammen. Das Sprachmodell gibt Wahrscheinlichkeiten für aufeinanderfolgende Wörter an. Die Wahrscheinlichkeiten des Sprachmodells werden beim Erkennungsprozess in folgender Approximation berücksichtigt:

$$\operatorname{argmax}_w p(w|x) \approx \operatorname{argmax}_w p(x|w) \cdot p(w)^{lz} \cdot |w|^{lp}$$

wobei $p(w|x)$ die Wahrscheinlichkeit des Auftretens eines Wortes nach einer gegebenen Wortfolge, $p(x|w)$ die Wahrscheinlichkeit der Beobachtung einer Wortfolge (Score) (angegeben als Gaussverteilung), $p(w)$ die Wahrscheinlichkeit einer Wortfolge (angegeben durch die Bi-, Tri- oder Viergramme im Sprachmodell) und $|w|$ die Länge der Wortfolge ist. lz und lp sind dabei Parameter, die während der Evaluierung ermittelt und dann dem Spracherkennner einprogrammiert werden müssen.

Die Ausgabe des Spracherkenners, dessen Grundlage das hier angeführte Sprachmodell ist, kann aus beliebig vielen Buchstaben bestehen. Es kann hierbei also weder gewährleistet werden, dass die Ausgabe immer aus vier Buchstaben besteht noch dass die Konventionen zum Aufbau der deutschen Flughafen-Codes eingehalten werden.

3.2 Spracherkennner mit N-Gramm-Modell (vollständige Codes)

Dem folgenden Sprachmodell liegt die Annahme zugrunde, dass durch das Zusammenfügen der einzelnen Buchstaben zu Flughafen-Codes bzw. die Benutzung der Flughafen-Codes als zu erkennende Wörter der Spracherkennner nur aus vier Buchstaben bestehende Ausgaben erzeugt und die Konventionen des Aufbaus der Flughafen-Codes eingehalten werden. Bei dem Spracherkennner auf vollständigen Codes wurden im Wörterbuch sowie im Vokabular also nicht einzelne Buchstaben angegeben, sondern es wurden durch Aneinanderhängen der jeweiligen Aussprache der Buchstaben, getrennt durch das Phonem für Stille, alle deutschen Flughafen-Codes zusammengefügt, wobei jeder Code mit zwei bis vierzig Aussprachevarianten im Wörterbuch verzeichnet ist.

Zum Beispiel:

EDBE {EH K OW SIL D EH L T AX SIL B R AH V OW SIL EH K OW}

EDBE(2) {EH K OW SIL D EH L T AA SIL B R AH V OW SIL EH K OW}

Das Wörterbuch und das Vokabular wurden nicht mehr von Hand erstellt, sondern es wurde mit Hilfe eines TCL-Skriptes aus dem Wörterbuch des Spracherkenners mit buchstabenbasiertem Sprachmodell generiert. Dabei wurde aus dem Wörterbuch ein Array mit den verschiedenen Aussprachevarianten der Buchstaben erstellt. Danach wurden nacheinander alle Flughafen-Codes aus der Datenbank ausgelesen, die Buchstaben der Flughafen-Codes durch die Aussprache dieser ersetzt, bei mehreren Aussprachevarianten für Buchstaben der Eintrag im Wörterbuch für den Flughafen-Code vervielfacht und die Aussprachen der Buchstaben durch das Phonem der Stille getrennt. Die so erhaltene Aussprachevariante für den Flughafen-Code wurde mit diesem in das Wörterbuch geschrieben und der Flughafen-Code mit der Nummerierung der jeweiligen Aussprachevariante im Vokabular vermerkt. Das Wörterbuch besteht somit aus 445 Flughafen-Codes mit bis zu 40 Aussprachevarianten pro Code und wenigen modellierten Störgeräuschen. Das Wörterbuch besitzt deshalb 4164 Einträge, was sich negativ auf die Performance des Spracherkennungsprozesses auswirkt.

Der beschriebene Aufbau ist praktikabel, sofern nur Flughafen-Codes zu erkennen sind und keine zusätzlich gesprochenen Worte. Da dies aber nicht der Fall ist, gilt dieser Spracherkennung von vorne herein für die Weiterverwendung in einem umfangreichen Flugnavigationssystem als unbrauchbar. Das Sprachmodell wurde wiederum mit dem Interactive Systems LM Toolkit (siehe [RiSG96]) basierend auf einer Trigrammstatistik aller deutschen Flughafen-Codes erstellt, wobei hauptsächlich Monogramme mit gleicher Wahrscheinlichkeit für alle Flughafen-Codes in diesem enthalten sind.

3.3 Spracherkennung mit CFG-Modell

Der Janus-Spracherkennung besitzt eine für den Zweck der Erkennung von Flughafen-Codes interessante Erweiterung. Durch diese Erweiterung ist es möglich, das Sprachmodell als kontextfreie Grammatik (CFG) anzugeben und während der Erkennungsphase zwischen verschiedenen Grammatiken zu wechseln. Der Aufbau des Wörterbuches und des Vokabulars ist genauso wie beim Spracherkennung mit N-Gramm-Modell (Buchstaben). Jedoch entfällt das Generieren eines Sprachmodells mit dem Interactive Systems LM Toolkit (siehe [RiSG96]) zugunsten der Angabe einer kontextfreien Grammatik. Da der Aufbau der Flughafen-Codes, wie in Kapitel 2.2 beschrieben, zwingend vorgegeben ist, legt dies die Verwendung einer kontextfreien Grammatik (eines endlichen Automaten) nahe.

In Tabelle 3.1 ist die verwendete Grammatik angegeben, wobei Nicht-Terminalsymbole groß und Terminal-Symbole klein geschrieben sind. Ein * vor einem Terminal-Symbol bedeutet dabei, dass das Terminal-Symbol beliebig oft (auch überhaupt nicht) auftreten darf, [alpha...zulu] steht als Platzhalter für einen ausgesprochenen Buchstaben aus dem Alphabet. Als Geräusche wurden bei allen Spracherkennern neben „Klick“- und Atemgeräuschen auch „äh's“ modelliert.

	→	AFSRA
AFSRA	→	NOISE echo STATE
STATE	→	NOISE delta AIS NOISE tango ARMY
AIS	→	NOISE alpha TOWN NOISE bravo TOWN NOISE charlie TOWN NOISE delta TOWN NOISE echo TOWN NOISE foxtrott TOWN NOISE golf TOWN NOISE hotel TOWN NOISE kilo TOWN NOISE lima TOWN NOISE mike TOWN NOISE november TOWN NOISE oscar TOWN NOISE papa TOWN NOISE quebec TOWN NOISE romeo TOWN NOISE sierra TOWN NOISE tango TOWN NOISE victor TOWN NOISE whisky TOWN NOISE xray TOWN NOISE zulu TOWN
ARMY	→	NOISE [alpha...zulu] TOWN
TOWN	→	NOISE [alpha...zulu] NOISE
NOISE	→	*click *breath *interjection *interjectionah

Tabelle 3.1: Modellierung des Aufbaus der Flughafen-Codes mit Hilfe einer kontext-freien Grammatik

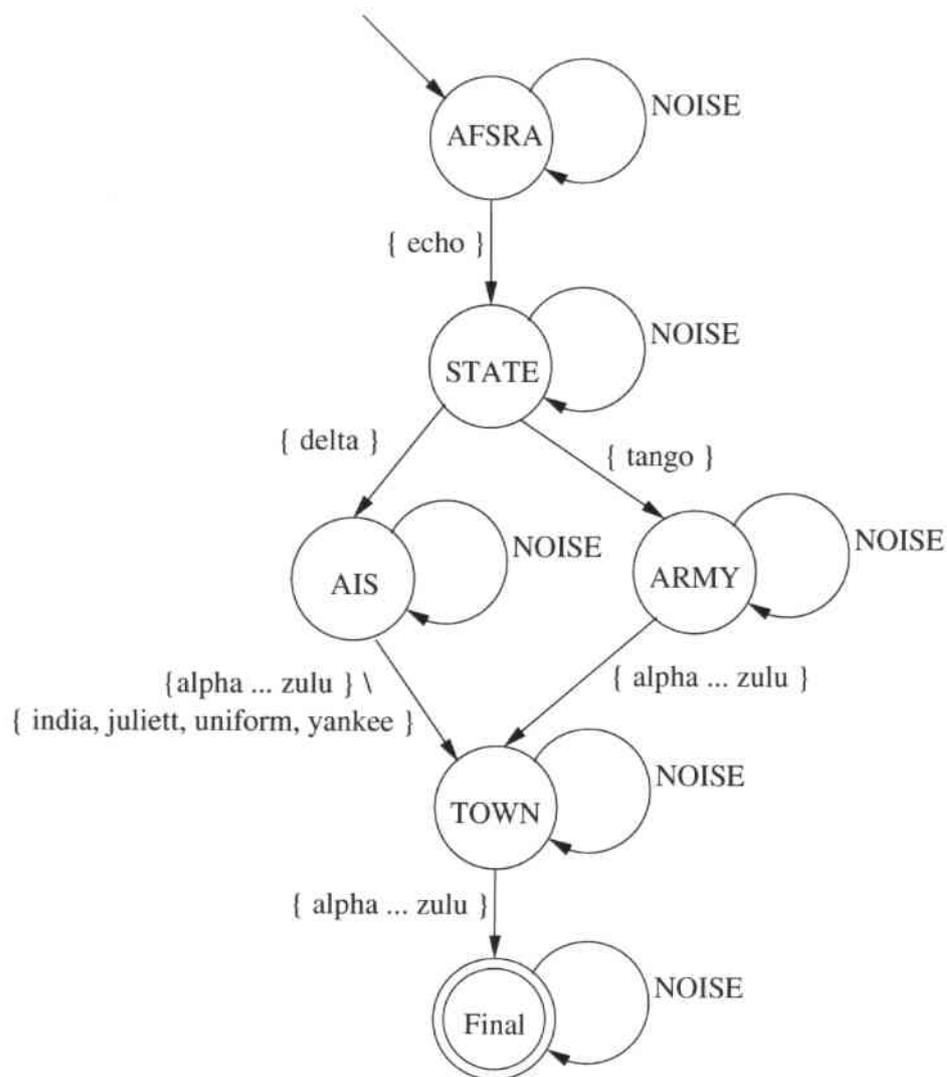


Abbildung 3.1: Flughafen-Codes erkennender Automat

Aus dieser Grammatik lässt sich ein endlicher Automat (Abb. 3.1) erstellen, der bei der Erkennung von gesprochenen Flughafen-Codes durchlaufen wird. Hierbei werden, besser als beim Spracherkenner mit N-Gramm-Modell (Buchstaben), die vorgegebenen Konventionen zum Aufbau der Codes eingehalten und es ist gewährleistet, dass jede Ausgabe des Spracherkenners aus maximal vier Buchstaben besteht.

4. Evaluierungsergebnisse

Zur Leistungsfeststellung der Spracherkennung mit unterschiedlichen Sprachmodellen wurde eine Evaluierung durchgeführt. Dies geschah mit 445 vom Verfasser dieser Arbeit gesprochenen Codes (alle deutschen Flughafen-Codes). Dies wurde für ausreichend erachtet, da der Verfasser weder an der Erstellung des Broadcast-News-Erkennung noch an dessen Training beteiligt war. Die Evaluierung sollte richtungsweisende Erkenntnisse für die weitere Erstellung des Flugnavigationssystems liefern.

4.1 Evaluierung mit geräuscharmen Sprachdaten

Wie oben bereits erwähnt, wurde die Evaluierung mit 445 gesprochenen Flughafen-Codes durchgeführt. Diese wurden mit einem mit einer Soundkarte mitgelieferten Mikrofon unbekannter Marke aufgenommen. Hierzu wurde ein Software-Tool geschrieben, das aus der Datenbank mit den Flughafen-Codes zufällig eine beliebige Anzahl an Codes auswählt und dem „Sprach-Spender“ präsentiert. Dieser kann dann mit einer „Ein-Knopf-Bedienung“ die Aufnahme des jeweiligen Codes selbständig starten und stoppen. Es wird sowohl der Flughafen-Code als auch dessen Aussprache dargestellt (siehe Abb. 4.1).

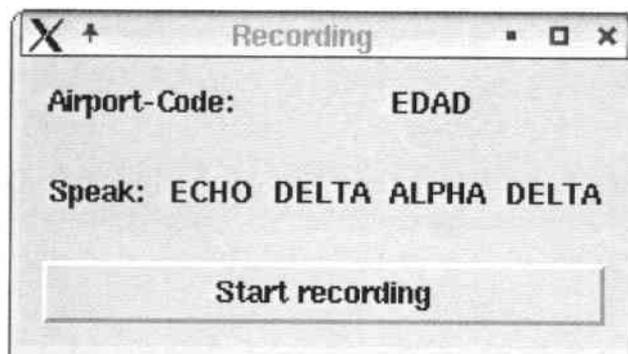


Abbildung 4.1: Software für die „Sprach-Spende“

Zur Evaluierung wurde ein weiteres TCL-Script geschrieben, welches die Sprachaufnahmen dem als eigenen Prozess laufenden Spracherkenner übergibt, die Spracherkennung startet und das Erkennungsergebnis auswertet. Der Datenaustausch zwischen dem Evaluierungsprogramm und dem Spracherkenner erfolgte dabei über das Anlegen von Dateien mit Hilfe des Unix-Befehls „touch“. Nach der Initialisierung und jedem abgeschlossenen Erkennungslauf legt der Spracherkenner eine Datei an, um damit die Bereitschaft zur Verarbeitung von Sprachaufnahmen zu signalisieren. Das Evaluierungsskript nimmt diese Datei zur Kenntnis, löscht die Datei, kopiert die jeweilige Sprachaufnahme in die entsprechende Eingabedatei des Spracherkenners und signalisiert dem Spracherkenner durch das Anlegen einer Datei, dass dieser mit der Erkennung beginnen soll. Hat der Spracherkenner die Sprachaufnahme ausgewertet, schreibt er das Ergebnis in eine Ausgabedatei und legt eine weitere Signalisierungsdatei an, um das Evaluierungsprogramm auf die bereitliegende Ausgabe hinzuweisen. Das Evaluierungsprogramm liest die Ausgabe und löscht die beide Dateien. Die Ausgabe wird nun mit der Transkription der Sprachaufnahme verglichen. Dabei wird zunächst festgestellt, ob die Ausgabe komplett der Transkription entspricht oder ob der „Satz“ falsch ist. Daraus wird am Ende der Evaluierung die Satzfehlerrate berechnet. Danach werden die einzelnen Worte überprüft, woraus später die Wortfehlerrate berechnet wird.

4.1.1 Spracherkenner mit N-Gramm-Modell (Buchstaben)

Bei diesem Spracherkenner wurden zahlreiche Testläufe absolviert. Zum einen um das trigrammbasierte Sprachmodell mit dem viergrammbasierten Sprachmodell vergleichen zu können. Zum anderen um geeignete Werte für die Parameter l_z und l_p zu erhalten. Insgesamt wurden dafür 37 Testläufe durchgeführt, die im Durchschnitt jeweils 1 Stunde 15 Minuten gedauert haben.

In Abbildung 4.2 sind die Ergebnisse der Testläufe mit dem Spracherkenner mit trigrammbasiertem Sprachmodell dargestellt. Der Broadcast-News-Erkennen arbeitete mit den Parametern $l_p = 10$ und $l_z = 26$. Ausgehend von diesen Werten wurde versucht, die Parameter in der „Umgebung“ dieser Werte zu variieren, um eine bessere Erkennungsrate zu erzielen. Dabei wurde der Parameter l_p gleich um zwei erhöht bzw. erniedrigt, da es nur auf signifikante Änderung der Satz- bzw. Wortfehlerrate ankam. Auf das Ausfüllen der gesamten Matrix zur Bestimmung des Optimums wurde verzichtet. Da der Parameter l_p nach der Approximation (siehe Kapitel 3.1)

$$\operatorname{argmax}_w p(w|x) \approx \operatorname{argmax}_w p(x|w) \cdot p(w)^{l_z} \cdot |w|^{l_p}$$

Einfluss auf die Satzlänge nimmt, bestand dabei nicht viel Optimierungsbedarf, da die Sätze bei allen Sprachdaten die Länge vier hatten. Der Parameter l_z gewichtet die Wahrscheinlichkeit einer Wortfolge und bestimmt damit die Wichtigkeit des Sprachmodells gegenüber der Akustik.

Eine graphische Darstellung der Wortfehlerraten der Matrix ist in Abbildung 4.2 zu sehen. Nach den Testläufen wurde das Minimum der Fehlerraten der Matrix gewählt. Die minimale Satzfehlerrate betrug 3,15% und die minimale Wortfehlerrate 1,12% bei den Werten 8 und 10 für den Parameter l_p und den Werten 26, 27 und 28 für

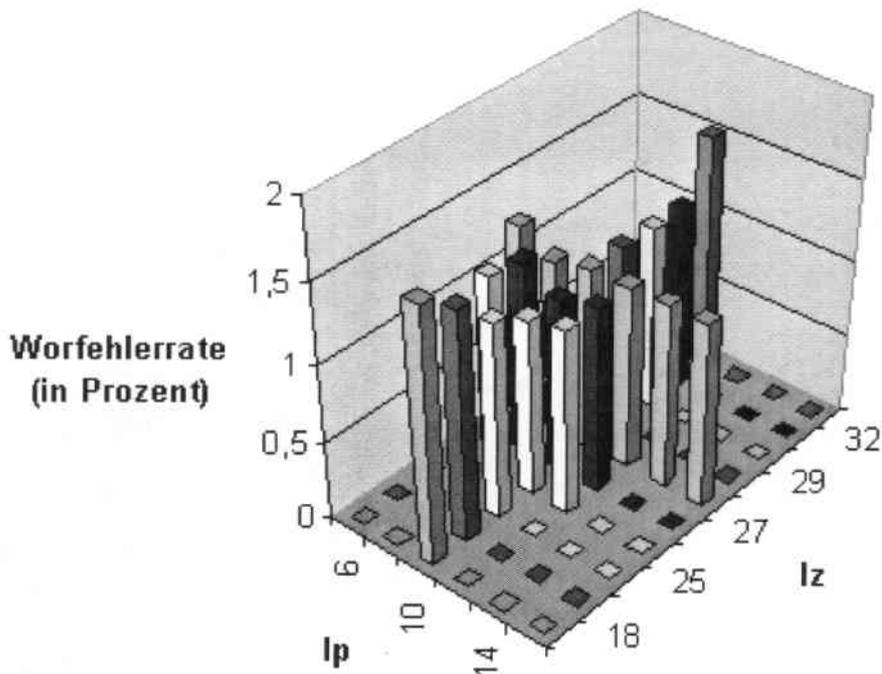


Abbildung 4.2: Evaluierungsergebnisse des Spracherkenners mit trigrammbasiertem Sprachmodell (Buchstaben)

den Parameter l_z . Aufgrund dieser Ergebnisse wurden für die Parameter l_p und l_z die Werte des Broadcast-News-Erkenners ($l_p = 10$, $l_z = 26$) übernommen.

Ähnlich wurden die Parameter für den Spracherkennung mit viergrammbasiertem Sprachmodell festgelegt. Die Ergebnisse des Spracherkenners mit viergrammbasiertem Sprachmodell (Abb. 4.3) stimmten fast mit den Ergebnissen des Spracherkenners mit trigrammbasiertem Sprachmodell (Abb. 4.2) überein. Dies ist im ersten Moment erstaunlich, da man von dem Spracherkennung mit viergrammbasiertem Sprachmodell eine bessere Erkennungsleistung erwartet hätte. Allerdings werden in den Viergrammen alle möglichen Kombinationen der Flughafen-Codes berücksichtigt, da das Sprachmodell auf der Basis aller deutschen Flughafen-Codes erstellt wurde. Dabei besitzen die Viergramme gleiche Wahrscheinlichkeiten, da das Auftreten eines jeden deutschen Flughafen-Codes gleichwahrscheinlich ist und die Viergramme dadurch nutzlos werden. Dies könnte eine Begründung für die gleichbleibende Erkennungsleistung sein.

Aus diesem Grund wurde die Wahl der Parameter des Spracherkenners mit trigrammbasiertem Sprachmodell übernommen und der Spracherkennung mit viergrammbasiertem Sprachmodell in der Evaluierung von geräuschbehafteten Daten ausgeschlossen.

4.1.2 Spracherkennung mit N-Gramm-Modell (vollständige Codes)

Die Ergebnisse der Evaluierung des Spracherkenners mit dem Sprachmodell der vollständigen Codes sind in Tabelle 4.1 wiedergegeben. Der Parameter l_p wurde bei der Evaluierung konstant gehalten und die Werte für den Parameter l_z wurden

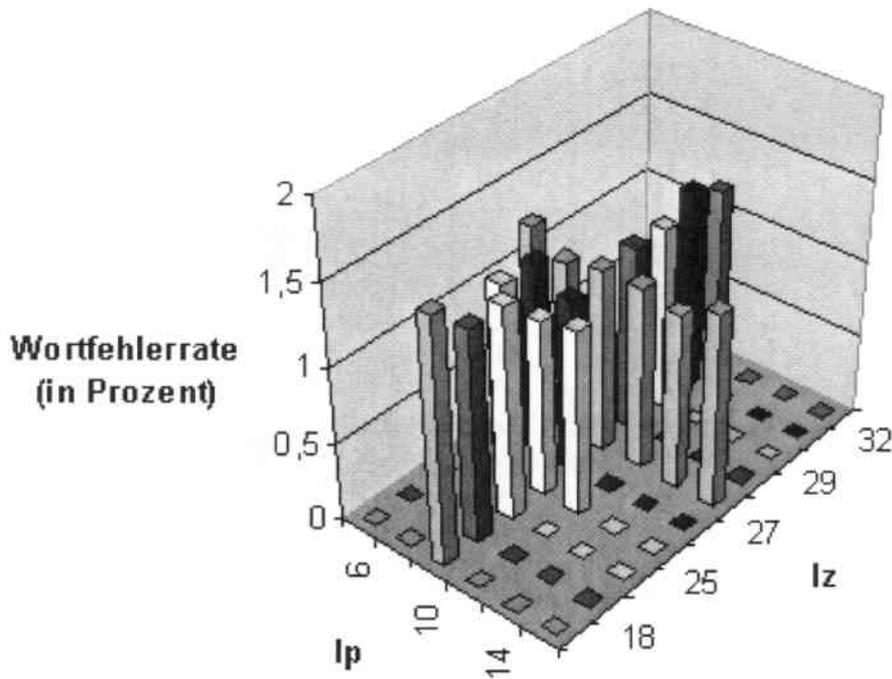


Abbildung 4.3: Evaluierungsergebnisse des Spracherkenners mit viergrammbasiertem Sprachmodell (Buchstaben)

lz	Satzfehlerrate	Wortfehlerrate
25	11,91%	3,15%
26	12,36%	3,26%
27	12,36%	3,31%
28	12,36%	3,31%
29	12,36%	3,31%
30	12,36%	3,31%

Tabelle 4.1: Evaluierungsergebnisse des Spracherkenners mit trigrammbasiertem Sprachmodell (vollständige Flughafen-Codes, $lp = 10$)

variiert. Die durchgehend schlechte Erkennungsleistung verwundert nicht, da im Sprachmodell, wie in Kapitel 3.2 festgestellt wurde, fast nur Monogramme zu finden sind, da das Sprachmodell auf der Basis der deutschen Flughafen-Codes modelliert wurde. In einem „Satz“ tritt dabei immer nur ein Flughafen-Code auf. Es folgt also auf einen Code nie ein zweiter. Da im Sprachmodell Wahrscheinlichkeiten für das Auftreten nachfolgender Wörter angegeben werden und es in einem Satz keine auf ein Wort folgenden Wörter gibt, werden alle Codes mit der gleichen Wahrscheinlichkeit gewichtet. Dieser Umstand macht sich dann in den Erkennungsergebnissen bemerkbar.

4.1.3 Spracherkenner mit CFG-Modell

Sehr erfreulich war das Ergebnis des Spracherkenners mit CFG-Modell. Hier mussten keine Parameter wie lz und lp bestimmt werden. Es genügte ein einziger Testlauf, der ein hervorragendes Ergebnis erbrachte, wie in Tabelle 4.2 zu sehen ist. Dies

Satzfehlerrate	Wortfehlerrate
0,45%	0,23%

Tabelle 4.2: Evaluierungsergebnis des Spracherkenners mit CFG-Modell

verwendete Grammatik	Wortfehlerrate
deutsche Flughafen-Codes	0,23%
deutsche und schweizer Flughafen-Codes	0,33%
deutsche, schweizer und österreichische Flughafen-Codes	0,50%

Tabelle 4.3: Evaluierung des Spracherkenners mit CFG-Modell mit unterschiedlichen Grössen der Grammatik

ist auf die Tatsache zurückzuführen, dass durch das Aufstellen der Grammatik die Satzlänge sowie die mögliche Abfolge der nach dem internationalen Buchstabieralphabet buchstabierten Buchstaben festgelegt ist. Der Spracherkennung muss sich beim Erkennungsprozess an diese Abfolge halten und erzielt damit die besten Ergebnisse.

Der Zusammenhang zwischen Wortfehlerrate und Größe der verwendeten Grammatik war beim Spracherkennung mit CFG-Modell zusätzlich interessant. Hierfür wurde die Grammatik in zwei Stufen erweitert, wobei das Wörterbuch und das Vokabular des Spracherkenners nicht verändert wurden. Zuerst wurde die Grammatik auf die zusätzliche Erkennung von Flughafen-Codes schweizer Flugplätze ergänzt, danach wurden die zur Erkennung von österreichischen Flughafen-Codes notwendigen Informationen hinzugefügt. Da die Flughafen-Codes, wie in Kapitel 2.2 beschrieben, weltweit (fast) einheitlich aufgebaut sind, ließ sich die Grammatik sehr einfach erweitern. Die Wortfehlerrate wurde bei dem Spracherkennung durch einen Testlauf pro Grammatik ermittelt, wobei jeder Testlauf mit 300 zufällig ausgewählten Testdaten durchgeführt wurde. Das Ergebnis ist in Tabelle 4.3 dargestellt. Es ist zu erkennen, dass die Wortfehlerrate mit steigender Grösse der Grammatik zunimmt, da der Spracherkennung im Rahmen des Erkennungsprozesses mehr Möglichkeiten zur Entscheidung für ein bestimmtes Phonem hat.

Testlauf	Satzfehlerrate	Wortfehlerrate
1	87,87%	51,63%
2	96,63%	73,31%
3	96,18%	73,54%

Tabelle 4.4: Evaluierungsergebnisse des Spracherkenners mit N-Gramm-Modell (Buchstaben) bei geräuschbehafteten Daten

4.2 Evaluierung mit geräuschbehafteten Sprachdaten

Zur Evaluierung der Spracherkennung mit geräuschbehafteten Daten wurden die aufgenommenen geräuscharmen Sprachdaten mit Helikoptergeräuschen in drei unterschiedlichen Einstellungen gemischt. Dabei wurde der Lombard-Effekt – die Veränderung der Aussprache und Betonung einer Person bei geräuschvollem Hintergrund – missachtet. Für den Testlauf 1 wurden die Sprachdaten mit den Helikoptergeräuschen ohne zusätzliche Regelung (Pegelanpassung, etc.) gemischt. Der Pegel des Helikoptergeräusches wurde dann sukzessive bis zum Testlauf 3 angehoben, so dass bei den Testdaten des dritten Testlaufes subjektiv das Helikoptergeräusch die Aufnahme dominiert, die Sprache aber von einem Menschen noch verstanden wird. Das Mischen der Geräusche mit den Sprachdaten wurde gegenüber Aufnahmen während eines Helikopterfluges vorgezogen, da eine Flugstunde des Helikopters sehr teuer ist und die Geräuschdaten in ausreichender Menge vorhanden waren.

Bei jedem Erkennen wurden drei Testläufe mit diesen unterschiedlichen Daten gemacht. Die Testläufe zwei und drei wurden in die Evaluierung aufgenommen, um die Notwendigkeit einer Geräuscheliminierung bzw. -unterdrückung vor der eigentlichen Spracherkennung darlegen zu können.

4.2.1 Spracherkennung mit N-Gramm-Modell (Buchstaben)

Der Spracherkennung mit N-Gramm-Modell (Buchstaben) wurde nur mit dem trigrammbasierten Sprachmodell evaluiert. Tabelle 4.4 zeigt die Ergebnisse – mit sehr schlechten Erkennungsraten, so dass für einen Einsatz im Helikopter, für den dieser Spracherkennung aufgrund der Erweiterungsmöglichkeiten grundsätzlich in Frage kommt, die Implementierung einer signalbasierten Geräuscheliminierung bzw. -unterdrückung mit Hilfe von zum Beispiel der Fouriertransformation oder Wavelets erfolgen muss.

Testlauf 3 wurde, um die Vermutung der Nutzlosigkeit der Viergramme im Sprachmodell zu bestätigen, auch mit dem viergrammbasierten Sprachmodell durchgeführt. Dabei wurde, wie erwartet, kein besseres Ergebnis erzielt. Die Satzfehlerrate erreichte 96,18%, die Wortfehlerrate 73,93%.

4.2.2 Spracherkennung mit N-Gramm-Modell (vollständige Codes)

Obwohl bereits als nicht praktikabel ausgeschlossen, wurde der Spracherkennung mit dem Sprachmodell basierend auf den vollständigen Flughafen-Codes mit den

Testlauf	Satzfehlerrate	Wortfehlerrate
1	37,30%	15,00%
2	59,33%	28,37%
3	62,47%	30,34%

Tabelle 4.5: Evaluierungsergebnisse des Spracherkenners mit N-Gramm-Modell (vollständige Flughafen-Codes) bei geräuschbehafteten Daten

Testlauf	Satzfehlerrate	Wortfehlerrate
1	30,79%	9,04%
2	53,26%	17,25%
3	54,16%	17,98%

Tabelle 4.6: Evaluierungsergebnisse des Spracherkenners mit CFG-Modell bei geräuschbehafteten Daten

Parametern $l_p = 10$ und $l_z = 26$ mit den geräuschbehafteten Daten getestet. Dabei zeigte dieser, wie in Tabelle 4.5 zu sehen, bessere Ergebnisse als der Spracherkennung mit einem Sprachmodell basierend auf einzelnen Buchstaben. Diese Ergebnisse wurden nicht erwartet.

4.2.3 Spracherkennung mit CFG-Modell

Der Spracherkennung mit CFG-Modell zeigte wiederum die besten Ergebnisse (siehe Tabelle 4.6), auch wenn diese sehr stark von den Ergebnissen der Evaluierung mit geräuscharmen Sprachdaten abweichen. Die Ergebnisse sind durch die Berücksichtigung der Konventionen des Aufbaus von Flughafen-Codes und die Festlegung der Satzlänge auf vier Buchstaben zu erklären. Trotzdem sollte auf jeden Fall eine signalbasierte Geräuscheliminierung bzw. -unterdrückung implementiert werden, um die Ergebnisse zu verbessern.

5. Zusammenfassung und Ausblick

Sprecherunabhängige sprachgesteuerte Flugnavigationssysteme in Helikoptern bzw. in der Allgemeinen Luftfahrt gibt es noch nicht. Dabei wären diese sehr von Vorteil, da dem Piloten Arbeit abgenommen und dieser dadurch entlastet wird. Nach einer Einführung in luftfahrtspezifische Grundlagen wurde in dieser Arbeit festgestellt, dass die für die Navigation in der Luftfahrt wesentlichen Ortskennungen für Flughäfen, hier Flughafen-Codes genannt, durch einen Spracherkenner bei geräuscharmen Sprachdaten mit geringer Fehlerrate erkannt werden. Der Spracherkenner mit CFG-Modell erbrachte bei der Evaluierung hervorragende Ergebnisse und kann daher für eine mögliche Implementierung des Flugnavigationssystems verwendet werden. Im Anhang A werden die im Rahmen dieser Arbeit programmierten Demonstrationen und deren Bedienung vorgestellt. Diese sind für die Verarbeitung von geräuscharmen Sprachdaten ausgelegt. Die Evaluierung geräuschbehafteter Sprachdaten hat gezeigt, dass für den Einsatz im Helikopter bzw. Flugzeug eine der Spracherkennung vorgeschaltete Geräuschunterdrückung bzw. -eliminierung erforderlich ist.

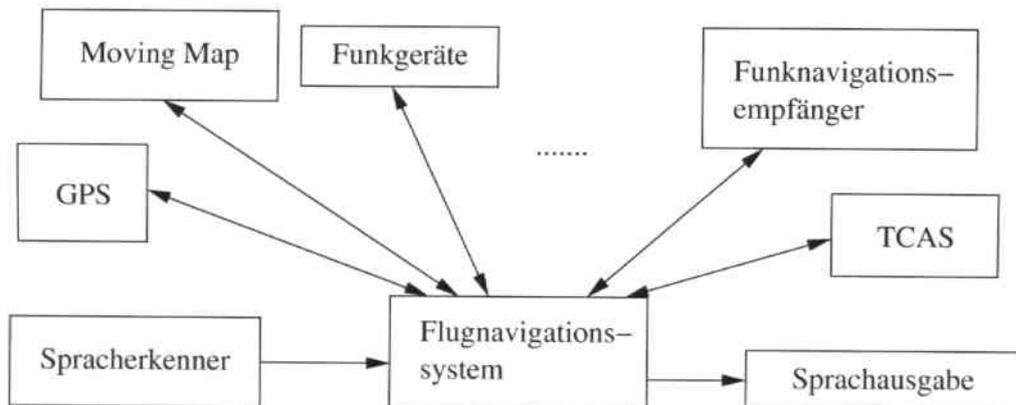


Abbildung 5.1: Interaktion des Flugnavigationssystems mit anderen Komponenten

Das Flugnavigationssystem soll mit einer Vielzahl von Geräten der Luftfahrt interagieren können. Die in Abbildung 5.1 dargestellten Komponenten¹ sind bis auf die Spracherkennung und -ausgabe kommerzielle Geräte, die auf dem entsprechenden Markt von einer Vielzahl von Herstellern angeboten werden und in fast allen Helikoptern bzw. Flugzeugen der Allgemeinen Luftfahrt vorhanden sind. Diese können über eine serielle oder andere Schnittstelle mit dem Flugnavigationssystem verbunden werden.

Nach Fertigstellung des Systems sollte der Pilot einen natürlichsprachlichen Dialog mit diesem führen können. Auch sollte es von zwei Piloten bedient bzw. benutzt werden können, da das Luftfahrzeug, je nach Typ, von zwei Piloten geflogen werden muss. Die Piloten arbeiten dann nach dem Crew Coordination Concept (CCC) zusammen und nehmen beide an verschiedenen Geräten des Helikopters bzw. des Flugzeuges unterschiedliche Einstellungen vor. Es sollte für die Piloten möglich sein, GPS-Routen (mit oder ohne Zuhilfenahme des internationalen Buchstabieralphabets) per Sprache einzugeben, Informationen über den Zielflugplatz und die aktuelle Position sowie aktuelle Daten des Luftfahrzeugs (Tankinhalt, Spritverbrauch, etc.) abzufragen und vieles mehr.

Aber nicht nur während, sondern auch vor dem Flug könnten neue Dienste bereitgestellt werden. Wird das System mit einem GSM-Mobiltelefon verbunden, so könnte der Pilot sich zum Beispiel vor dem Flug, wie in §3a Absatz (2) Luftverkehrsordnung (LuftVO) gefordert, über die aktuelle Wetterlage informieren oder einen Flugplan aufgeben.

Es gibt also viele Anwendungsbereiche im Cockpit eines Luftfahrzeuges, in denen Spracherkennung gekoppelt mit einem Flugnavigationssystem oder besser mit einem Flight-Management-System (FMS) eingesetzt werden kann.

¹TCAS – Traffic Alert and Collision Avoidance System

A. Aufbau und Benutzung der Demonstrationsversionen

Aus den aufgeführten und besprochenen Spracherkennern mit verschiedenen Sprachmodellen wurden zwei Demonstrationsversionen erstellt. Dabei wurde der Spracherkennung mit dem Sprachmodell basierend auf der Angabe der Flughafen-Codes als Wörter weggelassen, da er für die Praxis als nicht praktikabel erscheint.

Nach dem Entpacken der Datei „FlughafenCodeErkennungDemo.tar.bz2“ erhält man ein Verzeichnis gleichen Namens. In diesem Verzeichnis befinden sich die Dateien „demo“ und „demoCfg“. Mit dem Befehl „source demo“ wird die Demonstrationsversion des Spracherkenners mit dem N-Gramm-Modell (Buchstaben) gestartet. Dieser erkennt nur die in Kapitel 2.2 besprochenen Flughafen-Codes, falls diese nach dem internationalen Buchstabieralphabet (siehe Tabelle 2.2) buchstabiert werden. Dabei wird allerdings nur das trigrammbasierte Sprachmodell verwendet.

Der Befehl „source demoCfg“ startet die Demonstrationsversion des Spracherkenners, dessen Sprachmodell auf der in Kapitel 3.3 vorgestellten kontextfreien Grammatik bzw. auf dem endlichen Automaten basiert. Zu Demonstrationszwecken wurde diese Grammatik erweitert (siehe Tabelle A.1), um vor den Flughafen-Codes gesprochene Anweisungen erkennen zu können. Diese Anweisungen sind: „go to“, „go direct to“, „route to“ und „make a route to“. Nur diese vier Anweisungen können erkannt werden. Der Pilot muss sich deshalb vor dem Flug Kenntnisse über die zu gebrauchenden Anweisungsphrasen aneignen. Dieser Umstand ist gerade nicht erwünscht und wurde deshalb nur zu Demonstrationszwecken implementiert. Es zeigte sich jedoch, dass man die Grammatik, das Wörterbuch, das Vokabular und somit die Menge der durch den Spracherkennung erkannten Wörter einfach erweitern kann.

Beide Demonstrationsversionen unterscheiden sich nur im jeweils benutzten Sprachmodell und ihrem Aufruf („source demo“ bzw. „source demoCfg“). Danach geben sich beide Versionen in der Bedienung gleich. Nach dem Start erscheint das Initialisierungsfenster des Spracherkenners (Abb. A.1). In diesem Fenster kann man als einzige Aktion die Demonstration beenden.

Der Spracherkennung sowie die Steuerung werden als verschiedene Prozesse ausgeführt, wobei zuerst die Steuerung gestartet wird, welche dann wiederum den Sprach-

	→	COMMAND
	→	AFSRA
COMMAND	→	NOISE go NOISE to AFSRA NOISE go NOISE direct NOISE to AFSRA NOISE make NOISE a NOISE route NOISE to AFSRA NOISE route NOISE to AFSRA
AFSRA	→	NOISE echo STATE
STATE	→	NOISE delta AIS NOISE tango ARMY
AIS	→	NOISE alpha TOWN NOISE bravo TOWN NOISE charlie TOWN NOISE delta TOWN NOISE echo TOWN NOISE foxtrott TOWN NOISE golf TOWN NOISE hotel TOWN NOISE kilo TOWN NOISE lima TOWN NOISE mike TOWN NOISE november TOWN NOISE oscar TOWN NOISE papa TOWN NOISE quebec TOWN NOISE romeo TOWN NOISE sierra TOWN NOISE tango TOWN NOISE victor TOWN NOISE whisky TOWN NOISE xray TOWN NOISE zulu TOWN
ARMY	→	NOISE [alpha...zulu] TOWN
TOWN	→	NOISE [alpha...zulu] NOISE
NOISE	→	*click *breath *interjection *interjectionah

Tabelle A.1: Erweiterte kontextfreie Grammatik

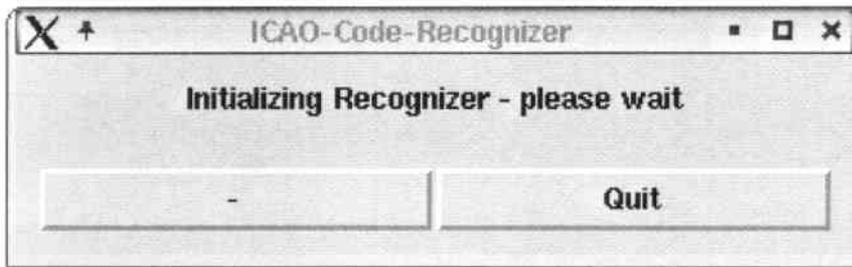


Abbildung A.1: Initialisierung

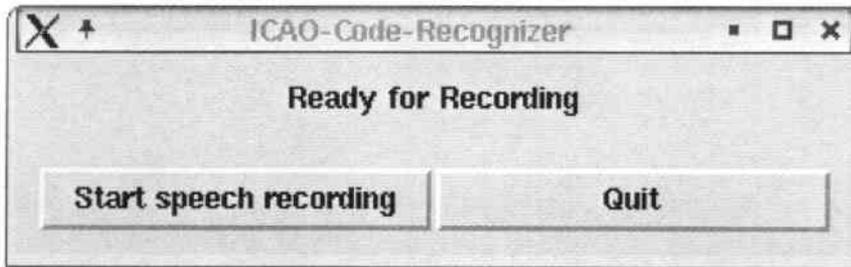


Abbildung A.2: Bereit zur Sprachaufnahme

erkenner aufruft. Vor dem Aufruf des Spracherkenners werden die jeweils benötigten Dateien für das Wörterbuch, das Vokabular und das Sprachmodell aus dem Unterverzeichnis „initial“ an die richtigen Stellen in der Verzeichnisstruktur der Demoversion kopiert. Eine Kommunikation zwischen Steuerung und Spracherkenner findet über das Erzeugen von Dateien mit Hilfe des Befehls „touch“ statt. So wird eine Datei vom Spracherkenner erzeugt, wenn dieser bereit ist, Sprachdaten zu verarbeiten. Die Ausgabe des Spracherkenners wird in eine Ausgabedatei geschrieben. Danach wird der Steuerung durch das Erzeugen einer Signalisierungsdatei mitgeteilt, dass die Ausgabe in die Ausgabedatei geschrieben wurde. Die Steuerung ihrerseits löscht nach Kenntnisnahme der Signalisierungsdateien diese. Die Ausgabe des Spracherkenners wird aus der Datei ausgelesen und es wird ein eigener Prozess gestartet, welcher die Ausgabe des Spracherkenners verarbeitet und alle in der Datenbank verfügbaren Daten in einem eigenen Fenster (Abb. A.5) darstellt.

Ist die Initialisierung des Spracherkenners abgeschlossen, wechselt die Beschriftung des linken Knopfes von „-“ auf „Start speech recording“ (Abb. A.2). Das System ist nun bereit, Sprache aufzunehmen. Dazu ist auf den linken Knopf zu klicken und anschliessend in das angeschlossene Mikrofon zu sprechen. Nach dem Drücken des Knopfes mit der Aufschrift „Start speech recording“ wechselt die Beschriftung auf „Stop speech recording“ (Abb. A.3). Somit kann nach erfolgter Spracheingabe ohne langes Suchen des entsprechenden Knopfes die Aufnahme beendet werden.

Nach dem Drücken des „Stop speech recording“-Knopfes wird die Aufnahme beendet, in das Verzeichnis „audio“ des Spracherkenners kopiert, zur Kontrolle wiedergegeben und die Spracherkennung gestartet (Abb. A.4). Nach Beendigung der Spracherkennung ist die Steuerung wieder bereit, Sprache aufzunehmen und dem Erkennen zuzuführen (Abb. A.2).

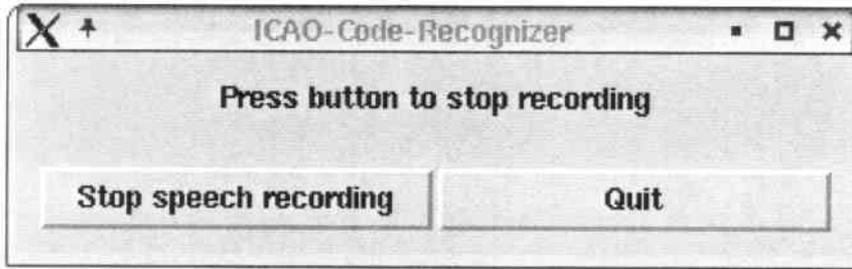


Abbildung A.3: Während der Sprachaufnahme

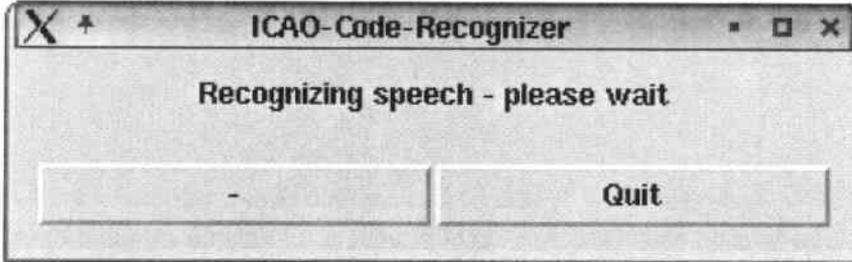


Abbildung A.4: Laufende Spracherkennung

Parallel dazu wird der Ausgabeprozess gestartet. Dieser stellt alle aus der Datenbank (siehe Kapitel 2.3, Tabelle 2.3) verfügbaren Informationen zu dem erkannten Flughafen-Code dar. In Abbildung A.5 ist das Ausgabefenster dargestellt. Über der Karte wird, allerdings nur bei dem Spracherkenner mit CFG-Modell, die eventuell gesprochene Anweisung dargestellt. In der Abbildung ist dies „GO DIRECT TO“. Darunter werden, falls verfügbar, eine Sichtanflugkarte des Flugplatzes und weitere Informationen dargestellt. Die Sichtanflugkarten wurden von der Internetseite <http://www.airports.de> heruntergeladen.

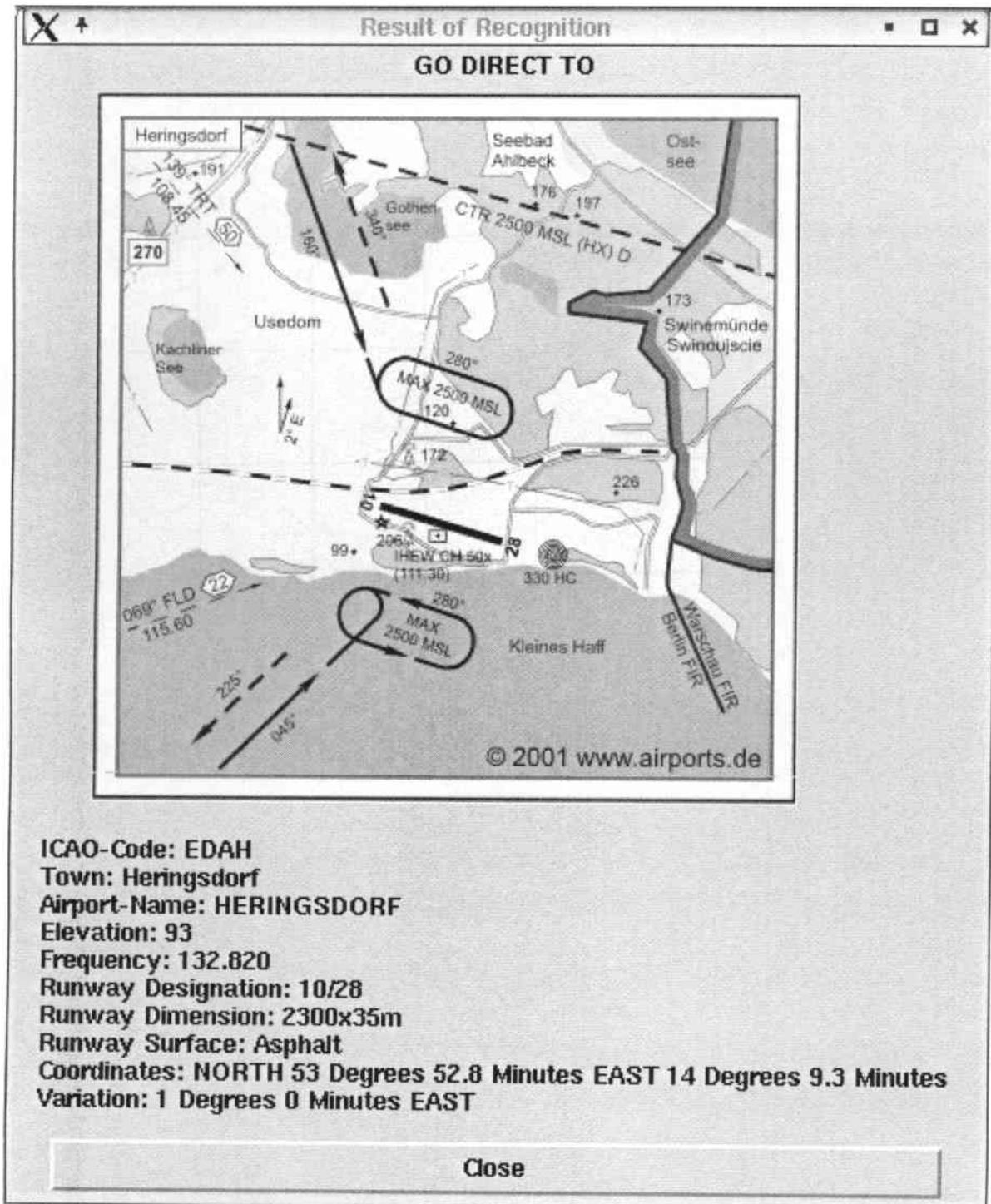


Abbildung A.5: Ausgabe der Spracherkennung

Literatur

- [airp01] Sichtanflugkarten deutscher Flugplätze. <http://www.airports.de>, 2001.
- [Albr02] Cathrin Albrecht. *Evaluierung multimodaler BOF*. Universität Hildesheim, Hauptseminar: Evaluierung in der Mensch-Maschine Interaktion, Dozent: Dr. T. Mandl. 2002.
- [Fran97] Dieter Franzen. *Kompaktlernprogramm zur Vorbereitung auf die Flugfunktionsprüfung AZF*. Bad Oeynhausener Flugfunk-Verlag. 1997.
- [Göts00] Ernst Götsch. *Luftfahrzeugtechnik*. Motorbuch-Verlag, Stuttgart. 2000.
- [KoSw] Robert Kobierski und Carl Swail. *Direct Voice Input for Control of an Avionics Management System*. National Research Council Ottawa, Canadian Marconi Company Kanata.
- [RiSG96] Klaus Ries, Bernhard Suhm und Petra Geutner. *Language Modeling in JANUS*. School of Computer Science, Carnegie Mellon University, Pittsburgh. 1996.
- [Rogi01] Dr. I. Rogina. *Spracherkennung-Praktikum*. Interactive Systems Laboratories, Universität Karlsruhe (TH). 2001.
- [Scha96] Thomas Schaaf. *Vertrauensmaße für die maschinelle Spracherkennung*. Institut für Logik, Komplexität und Deduktionssysteme, Universität Karlsruhe (TH). 1996.
- [Schu95] Ernst Günter Schukat-Talamazzini. *Automatische Spracherkennung - Statistische Verfahren der Musteranalyse*. Vieweg-Verlag. 1995.
- [Sibl98] Kathleen Sibley. *Speech system helps helicopter pilots stay safe*. Computing Canada. 1998.
- [WaRo95] A. Waibel und I. Rogina. *The Janus Speech Recognizer*. Interactive System Laboratories, Universität Karlsruhe (TH). 1995.
- [Wimm01] Thomas Wimmer. *Airports - Aktualisierte Excel-Datei mit den Deutschen Airports*. Nach ICAO, Name, etc. sortierbar. <http://www.eddh.de>, 2001.

- [WWFK⁺95] A. Waibel, M. Woszczyna, M. Finke, T. Kemp, A. McNair, A. Lavié, L. Mayfield, M. Maier, I. Rogina, T. Sloboda, P. Zahn und T. Zeppenfeld. *Janus II - Advances in spontaneous speech translation*. Interactive System Laboratories, Carnegie Mellon University and University of Karlsruhe (TH). 1995.