*TNO-report*
TM-96-A055

TNO Human Factors
Research Institute

title

# Development and performance of a cockpit control system operated by voice: summary report of project DMKLu/AC02/A/9105

Kampweg 5
P.O. Box 23
3769 ZG Soesterberg
The Netherlands

Phone +31 346 35 62 11
Fax +31 346 35 39 77

author
H.J.M. Steeneken

date
29 November 1996

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

All rights reserved.
No part of this publication may be reproduced and/or published by print, photoprint, microfilm or any other means without the previous written consent of TNO.

In case this report was drafted on instructions, the rights and obligations of contracting parties are subject to either the Standard Conditions for research instructions given to TNO, or the relevant agreement concluded between the contracting parties.
Submitting the report for inspection to parties who have a direct interest is permitted.

© 1996 TNO

number of pages : 13 (incl. appendices, excl. distribution list)

19970303 092

DTIC QUALITY INSPECTED 1

TNO Human Factors Research Institute
is part of TNO Defence research
to which also belong:
TNO Physics and Electronics Laboratory
TNO Prins Maurits Laboratory

Netherlands Organization for
Applied Scientific Research

# REPORT DOCUMENTATION PAGE

| | | |
|---|---|---|
| 1. DEFENCE REPORT NUMBER (MOD-NL)<br>RP 96-0198 | 2. RECIPIENT'S ACCESSION NUMBER | 3. PERFORMING ORGANIZATION REPORT NUMBER<br>TM-96-A055 |
| 4. PROJECT/TASK/WORK UNIT NO.<br>786.4 | 5. CONTRACT NUMBER<br>A91/KLu/325 | 6. REPORT DATE<br>29 November 1996 |
| 7. NUMBER OF PAGES<br>13 | 8. NUMBER OF REFERENCES<br>19 | 9. TYPE OF REPORT AND DATES COVERED<br>Final |

10. TITLE AND SUBTITLE

Development and performance of a cockpit control system operated by voice: summary report of project DMKLu/AC02/A/9105

11. AUTHOR(S)

H.J.M. Steeneken

12. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

TNO Human Factors Research Institute
Kampweg 5
3769 DE SOESTERBERG

13. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Director of Airforce Research and Development
Binckhorstlaan 135
2516 BA DEN HAAG

14. SUPPLEMENTARY NOTES

15. ABSTRACT (MAXIMUM 200 WORDS, 1044 BYTE)

Phase III of this project is concerned with the evaluation of an automatic speech recognizer for cockpit control functions in the MLU-F16. The results of this study are presented in three reports from which in this report an overview is given.
**Report III.0:** (TNO-TM), Summary report of Phase III (this report). **Report III.1:** (TNO-TM), Automatic speech recognition performance in a simulation-based fast-jet cockpit application. **Report III.2:** (TNO-TM), Spontaneous-speech data base for cockpit control applications applied to commercial state-of-the-art speech recognition technology. **Report III.3:** (NLR), Evaluation of integrated automatic speech recognition on the NSF mid-life update F-16 simulator.
A total of 29 sessions were flown during shake-down and training stages yielding 32.5 hours of recording. In 17 of these sorties three RNLAF pilots were participating.
The overall achieved word recognition accuracy was around 0.69, with scores per session ranging 0.53 to 0.88. The average completion rate (i.e. correctly executed commands) was around 66%. This is a low performance and insufficient for the envisaged operational applications.
The pilot debriefing information learned that although the performance was considered insufficient, the expansion of functions, such as radio station selection by name, was highly appreciated.
In general the present syntax was too complex which lead to incorrect commands. Also the awareness of the node status of the recognizer was marginal. A more flexible command language is an important requirement.
With the recorded speech signals of 17 sorties a data base was compiled. With this data base a repetition of the recognition experiments can be made with different types of recognizers.
Assessment of a new large vocabulary speech recognizer which was trained for the grammar (command string construction) of the cockpit commands produced a significantly higher recognition performance (0.87).

| | |
|---|---|
| 16. DESCRIPTORS<br>Automatic Speech Recognition<br>Cockpit<br>Command & Control | IDENTIFIERS |

| | | |
|---|---|---|
| 17a. SECURITY CLASSIFICATION (OF REPORT) | 17b. SECURITY CLASSIFICATION (OF PAGE) | 17c. SECURITY CLASSIFICATION (OF ABSTRACT) |
| 18. DISTRIBUTION/AVAILABILITY STATEMENT<br>Unlimited availability | | 17d. SECURITY CLASSIFICATION (OF TITLES) |

| | | |
|---|---|---|
| titel | : | Ontwikkeling en prestatie van een door spraak bediend cockpit-besturingssysteem; samenvattingsrapportage van fase 3 van het project DMKLu/AC02/A/9105 |
| auteur | : | H.J.M. Steeneken |
| datum | : | 29 november 1996 |
| opdrachtnr. | : | A91/KLu/325 |
| IWP-nr. | : | 786.4 |
| rapportnr. | : | TM-96-A055 |

In fase III van dit project werd de evaluatie uitgevoerd van een automatische spraakherkenner waarmee controlefuncties in de MLU-F16 cockpit konden worden uitgevoerd. De resultaten van deze evaluatie zijn samengevat in drie rapporten waarvan in dit rapport een overzicht wordt gegeven.

Rapport III.0: (TNO-TM), Summary report of Phase III (dit rapport).
Rapport III.1: (TNO-TM), Automatic speech recognition performance in a simulation-based fast-jet cockpit application.
Rapport III.2: (TNO-TM), Spontaneous-speech data base for cockpit control applications applied to commercial state-of-the-art speech recognition technology.
Rapport III.3: (NLR), Evaluation of integrated automatic speech recognition on the NSF mid-life update F-16 simulator.

In totaal werden 29 evaluatie sessies uitgevoerd die tezamen 32.5 uur aan opnamen opleverden. Hiervan werden 17 sorties uitgevoerd door drie KLu vliegers waarvan de resultaten werden geanalyseerd. De gemiddelde prestaties van de spraakherkenner bleek slechts 0,69 (accuracy) te bedragen met waarden tussen 0,53 en 0,88. Het gemiddelde percentage juist uitgevoerde commando's bedroeg 66%. Dit is marginaal en onvoldoende bij gebruik in operationele situaties, waartoe een percentage van 95% als ondergrens wordt beschouwd.
Bij de debriefing van de vliegers werd vastgesteld dat, ofschoon de prestaties ver onder de maat waren, de uitbreiding van de controlefuncties zeer werd gewaardeerd. Een voorbeeld hiervan is het selecteren van de radiofrequentie met de naam van de basis in plaats van de frequentie.
Over het algemeen was de gebruikte syntax van de commando's te ingewikkeld. Dit leidde tot verkeerde commando's waarvoor de herkenner niet was ingesteld. Ook was de vlieger zich niet altijd bewust van de status van de herkenner zodat commando's werden gegeven die niet in de syntax node voorkwamen en derhalve niet werden herkend. Een flexibele commandostructuur is dan ook een vereiste bij verdere toepassing.
Met de spraaksignalen die werden opgenomen gedurende de 17 sorties werd een spraakbestand samengesteld. Hiermee is het mogelijk andere herkenners te evalueren.
De evaluatie van een recent beschikbaar gekomen herkenner die geschikt is voor een grote vocabulaire en werd getraind met de grammatica van de gebruikte commando's (domein specifieke taal) leverde een gemiddelde score van 0,87.

CONTENTS Page

Report No.: TM-96-A055

Title: Development and performance of a cockpit control system operated by voice: summary report of project DMKLu/AC02/A/9105

Author: Dr. ing. H.J.M. Steeneken

Institute: TNO Human Factors Research Institute
Group: Perception

Date: November 1996

DO Assignment No.: A91/KLu/325

No. in Program of Work: 786.4

---

SUMMARY

Phase III of this project is concerned with the evaluation of an automatic speech recognizer for cockpit control functions in the MLU-F16. The results of this study are presented in three reports from which in this report an overview is given.

| | |
|---|---|
| Report III.0: | (TNO-TM), Summary report of Phase III (this report). |
| Report III.1: | (TNO-TM), Automatic speech recognition performance in a simulation-based fast-jet cockpit application. |
| Report III.2: | (TNO-TM), Spontaneous-speech data base for cockpit control applications applied to commercial state-of-the-art speech recognition technology. |
| Report III.3: | (NLR), Evaluation of integrated automatic speech recognition on the NSF mid-life update F-16 simulator. |

A total of 29 sessions were flown during shake-down and training stages yielding 32.5 hours of recording. In 17 of these sorties three RNLAF pilots were participating.
The overall achieved word recognition accuracy was around 0.69, with scores per session ranging 0.53 to 0.88. The average completion rate (i.e., correctly executed commands) was around 66%. This is a marginal performance and insufficient for the envisaged operational applications.

The pilot debriefing information learned that although the performance was considered insufficient, the expansion of functions, such as radio station selection by name, was highly appreciated.
In general the present syntax was too complex which lead to incorrect commands. Also the awareness of the node status of the recognizer was marginal. A more flexible command language is an important requirement.

With the recorded speech signals of 17 sorties a data base was compiled. With this data base a repetition of the recognition experiments can be made with different types of recognizers.
Assessment of a new large vocabulary speech recognizer which was trained for the grammar (command string construction) of the cockpit commands produced a significantly higher recognition performance (0.87).

Rap.nr. TM-96-A055 TNO Technische Menskunde
Soesterberg

**Ontwikkeling en prestatie van een door spraak bediend cockpit-besturingssysteem; samenvattingsrapportage van fase 3 van het project DMKLu/AC02/A/9105**

H.J.M. Steeneken

SAMENVATTING

In fase III van dit project werd de evaluatie uitgevoerd van een automatische spraakherkenner waarmee controlefuncties in de MLU-F16 cockpit konden worden uitgevoerd. De resultaten van deze evaluatie zijn samengevat in drie rapporten waarvan in dit rapport een overzicht wordt gegeven.

| | |
|---|---|
| Rapport III.0: | (TNO-TM), Summary report of Phase III (dit rapport). |
| Rapport III.1: | (TNO-TM), Automatic speech recognition performance in a simulation-based fast-jet cockpit application. |
| Rapport III.2: | (TNO-TM), Spontaneous-speech data base for cockpit control applications applied to commercial state-of-the-art speech recognition technology. |
| Rapport III.3: | (NLR), Evaluation of integrated automatic speech recognition on the NSF mid-life update F-16 simulator. |

In totaal werden 29 evaluatie sessies uitgevoerd die tezamen 32.5 uur aan opnamen opleverden. Hiervan werden 17 sorties uitgevoerd door drie KLu vliegers waarvan de resultaten werden geanalyseerd. De gemiddelde prestaties van de spraakherkenner bleek slechts 0,69 (accuracy) te bedragen met waarden tussen 0,53 en 0,88. Het gemiddelde percentage juist uitgevoerde commando's bedroeg 66%. Dit is marginaal en onvoldoende bij gebruik in operationele situaties, waartoe een percentage van 95% als ondergrens wordt beschouwd.
Bij de debriefing van de vliegers werd vastgesteld dat, ofschoon de prestaties ver onder de maat waren, de uitbreiding van de controlefuncties zeer werd gewaardeerd. Een voorbeeld hiervan is het selecteren van de radiofrequentie met de naam van de basis in plaats van de frequentie.
Over het algemeen was de gebruikte syntax van de commando's te ingewikkeld. Dit leidde tot verkeerde commando's waarvoor de herkenner niet was ingesteld. Ook was de vlieger zich niet altijd bewust van de status van de herkenner zodat commando's werden gegeven die niet in de syntax node voorkwamen en derhalve niet werden herkend. Een flexibele commandostructuur is dan ook een vereiste bij verdere toepassing.
Met de spraaksignalen die werden opgenomen gedurende de 17 sorties werd een spraakbestand samengesteld. Hiermee is het mogelijk andere herkenners te evalueren.
De evaluatie van een recent beschikbaar gekomen herkenner die geschikt is voor een grote vocabulaire en werd getraind met de grammatica van de gebruikte commando's (domein specifieke taal) leverde een gemiddelde score van 0,87.

# 1 INTRODUCTION

In 1991 a National Technology Project was started on the application of automatic speech recognition for system control in a fast jet cockpit. Two laboratories participated in this project: the TNO Human Factors Research Institute (TNO-TM, main contractor) and the National Aerospace Laboratory (NLR, sub-contractor). The project was divided in three phases:

Phase I: Laboratory evaluation of speech recognition systems, and identification of airborne applications (completed July 1992).

Phase II: Experimental assessment of the effectiveness of the application of speech recognition technology in a simulator environment (completed May 1995).

Phase III: Implementation and evaluation of a system in the National Simulator Facility (NSF) test-bed (originally in an aircraft). Completed October 1996.

The results obtained in Phase I were reported in five deliverables:

Report I.0: (TNO-TM), Summary report of Phase I,

Report I.1: (TNO-TM), Literature review on the state-of-the-art of automatic speech recognition,

Report I.2: (NLR), Literature review on airborne applications of automatic speech recognition,

Report I.3: (TNO-TM), Laboratory evaluation of five automatic speech recognition systems,

Report I.4: (NLR), Identification of automatic speech recognition applications in the F-16 MLU cockpit.

The results obtained in Phase II were reported in four deliverables.

Report II.0: (TNO-TM), Summary report of Phase II.

Report II.1: (TNO-TM), Development and assessment of the electro-acoustical input environment for automatic speech recognition in the cockpit.

Report II.2: (TNO-TM), Prediction of the performance of automatic speech recognition in a fast jet cockpit.

Report II.3: (NLR), Development and simulator implementation of an automatic speech recognition application for the mid-life update F-16 Cockpit.

The results obtained in Phase III are reported in four deliverables.

Report III.0: (TNO-TM), Summary report of Phase III (this report).

Report III.1: (TNO-TM), Automatic speech recognition performance in a simulation-based fast-jet cockpit application.

Report III.2: (TNO-TM), Spontaneous-speech data base for cockpit applications and assessment of commercial state-of-the-art speech recognition technology.

Report III.3: (NLR), Evaluation of integrated automatic speech recognition on the NSF Mid-life update F-16 simulator.

## 2 SUMMARY OF THE REPORTS III.1 THROUGH III.3

Report III.1 (TNO-TM): **Automatic speech recognition performance in a simulation based fast-jet cockpit application.**

The experiments consisted of 29 sorties of approximately one hour each. For 17 of these sorties three pilots of the RNLAF participated, the results of these sorties were analyzed. During each sortie the pilot in the F-16 National Simulator Facility had access to a control by voice of radio systems, displays and HOTAS functions (hands-on-throttle-and-stick). These systems could also be controlled manually as in the normal situation. During the "flight" tests recordings were made of the speech signals and a video recording of the pilot actions.

Analysis of all pilot actions including the voice control and debriefing was performed by the NLR and is reported separately. In this report the recognizer performance is analyzed. It was found that under these simulator flight conditions the performance (accuracy) drops from over 0.95 for read speech to 0.69 for the simulator spontaneous speech condition. Results obtained in four flight experiments performed in other laboratories showed similar results for read speech (three experiments) and for spontaneous speech (one experiment).

Analysis of the words used in the command strings showed that from the original 281 word vocabulary only 65 words were used frequently. These 65 words had a coverage of 90% of all words used during the tests. This means that the complexity of the recognition process can be reduced which will lead to a better performance of the recognizer.

From the speech material a calibrated date base was built with all the speech utterances annotated orthographically at command string level (described in the next section).

A pilot study was performed with a modern phoneme/grammar based recognizer. With this speaker independent system a mean performance of 0.85 (accuracy) was obtained. It is expected that this performance will exceed the 0.95 if this type of recognizer is trained for the non-native English speaking pilots rather than for, presumably read, American English speech. Also training with more representative speech signals obtained through an oxygen mask is required. It is foreseen that we will perform experiments with such a system in the near future.

Report III.2 (TNO-TM): **Spontaneous-speech data base for cockpit control applications applied to commercial state-of-the-art speech recognition technology.**

Additionally to a study on the performance of voice control of cockpit functions in a fast-jet simulator, a data base was made with the speech utterances of the pilots who participated in the experiments. With this data base an assessment was performed with two state-of-the-art large vocabulary recognizers.

*Data base recordings*

The voice input experiments in the MLU-F16 simulator produced a data base of 17 sorties of approximately one hour each. The environmental condition of the sorties were that a pilot operated a simulator aircraft and had control of the radio, display, and HOTAS functions both manually and by voice. The speech commands are spoken spontaneously and normally performed simultaneously with other tasks (flying the aircraft). For voice commands a special push-button was installed (PTT action). Around 29 sorties were flown from which 17 were selected to be included in the data base. These 17 sorties are flown by three experienced pilots (no experience with automatic speech recognition). During the experiments the equipment (radio, display, and HOTAS) were controlled both by the recognizer and manually. If the recognizer did not respond correctly the pilots tried normally again. Hence a realistic data base was obtained. All the speech utterances were annotated orthographically (text in ASCII characters). Additionally the PTT actions were recorded. This allows for separation of real commands from conversational speech. The data base recordings are available on five CD-ROM's according to the standard NIST format. Therefore, possible repetition of the experiments with present state-of the-art systems is relatively easy.
The data base described here is quite representative for pilot actions. Recently data bases were recorded elsewhere in flying aircraft. However, all these data bases are concerning read speech spoken by a co-pilot or navigator and not by the pilot who is flying the aircraft.

*State-of-the-art systems*

An overview of commercial state-of-the-art speech recognizers shows that the majority of present technology offers the following features: speaker independent, trained at the factory for American English, handling of a large vocabulary (> 20,000 words), and making use of a specific grammar which can be trained for a specific domain (e.g., cockpit control). In principle these systems can also be trained with specific speech signals supplied by a customer but this facility is not yet offered. Hence, the specific "oxygen mask speech" which defines the speech quality cannot be included at this moment. However, the grammar of the control words can be trained easily with some of the recognizers. We experimented with the IBM Voice type application factory which allows training of a grammar. It was shown that training with commands strings of half of the sorties (i.e., 8 sorties) and testing with the speech utterances of the other 9 sorties resulted in an accuracy for the three pilots of respectively: 0.74, 0.90, and 0.90. Hence a significant improvement was obtained. As these tests were performed with the factory trained system, it is expected that training with the (non-native English) pilots, and speaking through an oxygen mask, may improve the scores to an accuracy level of 0.95.

Report III.3 (NLR): **Evaluation of integrated automatic speech recognition on the NSF mid-life update F-16 simulator.**

The Phase III operational evaluations of the ASR applications on the F-16 simulator were executed from mid 1995 until mid 1996. The primary aim was the assessment of the effects on pilot attention distribution as a result of improved head-out and hands-on capabilities.

More than 20 sessions were flown during shake-down and training stages. A total of 29 evaluation sessions were flown, yielding 32.5 hours of recordings. In the various stages 6 RNLAF pilots/engineers participated; 3 test pilots and 2 NLR project members were available to participate in the actual evaluations.

The overall achieved word recognition rate (accuracy) was around 0.75with per session ranging from 0.57 to near 0.90. The average completion rate (i.e., correctly executed control actions) was around 66%. This marginal performance is not only apparent in the data results; it also affected the intended integration concept, the evaluation execution and the pilot comments.

The achievable recognition/completion rates for the tested system are judged insufficient for the envisaged operational applications. With the current state-of-the-art the operational potential of ASR is low. The objective to improve pilot attention distribution can not be achieved.
However, good sight was obtained in the operational requirements and in the functions that are eligible for "voice" applications. Functions which mainly mimic button action (especially for the current F-16 HOTAS switches) are not appreciated. Pilots favour functions which expand on the existing controls. Future ASR applications should aim for the realization of only a small number of functions with new selection possibilities (e.g., radio station selection by name), improved (direct) access and with Crew-Assistant like aspects (e.g., interactive checklists). Pilot requirements for direct access and intelligent (i.e., operational context sensitive) response dictate ASR systems to be fully integrated in the aircraft avionics, as opposed to the "add-on" concept adopted for the project.

Based on the five-year hands-on project experience, various recommendations are made with respect to required ASR technology developments while interfacing capabilities for future recognizers are stipulated. Furthermore, expected consequences of future ASR integration at aircraft/avionics level are identified.

Concerning recommendations for future R&D into ASR applications in a fighter cockpit environment, NLR would advise to await the availability of recognizers with proven better recognition capabilities. However, current state-of-the-art laboratory tests alone are not sufficient to prove all operationally relevant performance aspects. Until reliable laboratory test methods are developed, actual hands-on application evaluations remain necessary.

Reviewing the pilot comments, it is felt that the pilots have a need for functional enhancement and crew support functions; they have less need for ASR technology as an alternative for current controls. With the state-of-the-art of fighter designs and the limited availability of mechanical control options, the introduction of such functions will, in turn, result in control requirements which may be easier met by the application of ASR.

Therefore, NLR refrains from recommending an immediate follow-up project. It is however recommended that the RNLAF/MOD reviews ongoing or planned R&D projects related to crew support functions, and opens the possibilities for these projects to include the investigations of the use of ASR as a possible means of control.

## 3 OTHER RESEARCH ON SPEECH RECOGNITION IN THE COCKPIT

During the course of this project various papers were presented at workshops and conferences in order to discuss the results with other researchers (Steeneken & Van Velden, 1993; Steeneken & Pijpers, 1996).

Automatic speech recognition in the cockpit of fast jet aircraft and helicopters is a subject studied in various nations. In the western countries studies are in progress in Canada, Germany, France, the Netherlands, the United Kingdom, and the USA. It is foreseen that the European Fighter Aircraft (EFA, Germany, Spain, UK) will have an ASR facility in the cockpit.
For space applications the European Space Agency (ESA) prepares an Advanced Crew Terminal (ACT) equipped with a voice input and output facility to be tested in space in 1997.

The study by Williamson (1996), and South (1996) were mainly focused on the effect of g-force on the performance of ASR systems. In both studies g-force up to 6g was included, both in an aircraft and in a centrifuge. The studies made use of read speech obtained from a co-pilot or a navigator. Both studies used connected digits (radio frequencies) and some other control words. In general it was found that g-force up to 3g has no major effect on the recognition performance. A significant difference of the effect of g-force was found between experienced and unexperienced aircrew.

Cordonnier (1996) performed tests with two types of command strings: setting the radio frequency with connected digits and control of a display engine. There was real control of the radio system but the display control was artificial. The speech data were recorded for later evaluation with a specially designed connected word recognizer. The performance of the system increased during separate tests of the speaker from a rate of 89% correct to 98% correct. The test was performed with trained speakers.

Prévôt and Onken (1995) used a connected speech recognizer (MR8) for control of an on-board pilot assistance system. The system was evaluated during simulator and real flight experiments. The recognizer performance (percentage correct commands) improved during the tests from approx 63% to 86%. The experiments were performed in a standard aircraft, hence no oxygen mask was used.

All these studies make use of connected word recognizers. The more recently developed phoneme/grammar based systems were only included in a pilot experiment in this study.

## 4 GENERAL RESULTS AND FUTURE VIEWS

Phase III concerns the evaluation of the implemented recognition system and control software in the National Simulator Facility test-bed. The original goal was to study the effect of voice control on the workload of the pilot. However, due to the availability of an insufficient number of trained pilots and the poor performance of the recognition system this goal could not be reached. It was decided to perform the experiments with only three pilots. In total 29 sorties were conducted including flights by technicians and NLR engineers. From 17 sorties, performed by pilots only, the results were evaluated. In general the recognition performance during these flights was much lower than during the laboratory evaluations. Also a significant difference in performance was found between the three pilots. The performance measure (accuracy) ranged from 0.60 to 0.81 (mean 0.69). By omitting the effect on the performance of speaking errors, the use of out of vocabulary words, hesitations, and push-to-talk errors, the performance could be improved and a mean accuracy of 0.78 was obtained. This is still far below the score of 0.95 which is considered to be a lower limit for satisfactory performance of a command and control system operated by voice.

A major difference between the laboratory tests and the flight simulator tests concerned the type of speech used for the assessment. In the laboratory conditions *read* speech (commands and words) were used, while the flight tests were based on *spontaneous* speech, sometimes uttered simultaneously with other actions. This is quite different from most of the tests performed in other studies (see section 3) and can be considered as an important reason for the reduced recognition performance.

It was found that the structure of the command sequences was too complex for the pilots, sometimes the pilots had no notice of the node status of the recognizer and consequently produced commands which were not recognized correctly. A more flexible command syntax is therefore required. Such a feature is feasible with the present ASR systems which are in general focused on a large vocabulary, but adjusted for application in a specific domain (e.g., cockpit control commands).

It was also found that the pilots frequently used 65 words of the 281 word vocabulary. Such a reduction of the effective vocabulary size will lead to a better recognizer performance if the recognizer is trained with this limited set of words.

The speech data recorded during the 17 sorties were compiled to a calibrated data base from which both the wave-form of the speech signals (16 bit, 16 kHz sample rate) and the orthographic transcription is stored. Assessment of a present state-of-the-art recognizer with this data base resulted into a mean accuracy of 0.86. It should be noticed that this recognizer was speaker independent and factory trained with non representative speech signals (i.e., not recorded inside an oxygen mask but with a high quality microphone). It is expected the performance goal of 0.95 accuracy can be reached if the commercial systems are delivered with a user training option.

Pilots favour functions which expand on the existing controls. Future ASR applications should aim for the realization of only a small number of functions with new selection possibilities (e.g., radio station selection by name), improved (direct) access and with Crew-Assistant like aspects (e.g., interactive checklists). Pilot requirements for direct access and intelligent (i.e., operational context sensitive) response dictate ASR systems to be fully integrated in the aircraft avionics, as opposed to the "add-on" concept adopted for the project.

In at least five countries research is in progress on the application of voice control in cockpits and space stations.

## ACKNOWLEDGEMENTS

The project started in 1991 as an initiative from the Scientific Research Section of the RNLAF, and the two participating laboratories TNO-TM and NLR. Much support was obtained by various project officers of the RNLAF in particular Maj. ir. Th. W.G. de Laat, Maj. ir. R. Hol, Ir. S.J.J. de Bruin, and Maj. ir. J.H.J. Verhulst which was very much appreciated.
Also many pilots participated in the experiments both in earlier stages for data-base recording and in the final experiments for performing simulator flight tests. As the test results from individual pilots are presented in the various results we restrict ourselves by thanking them anonymously.

REFERENCES

Cordonnier, A., Kientz, D. & Wassner, H. (1996). Commande Vocale sur Aeronefs Militaires: Evaluation en Vol et Resultats. *Proc. AGARD Symposium "Audio effectiveness in aviation"*, Copenhagen (in press).

Eertink, B.J. & Pijpers, E.W. (1996). *Evaluation of integrated automatic speech recognition on the NSF Mid-life update F-16 simulator* (Report NLR CR 96646 L **III.4** Volume I and II). Amsterdam: National Aerospace Laboratory (NLR).

Houtgast, T. & Van Velden, J.G. (1992). *Application of automatic speech recognition in the fighter cockpit. Summary report on phase I* (Report IZF 1992 A-24, DMKLu/AC02/A/9105-report **I.0**). Soesterberg: TNO Institute for Perception.

Pijpers, E.W. & Van deGraaff, R.C. (1992). *Identification of automatic speech recognition applications in the F-16 Mid Life Update cockpit* (Report NLR CR 92 336 L **I.4**). Amsterdam: National Aerospace Laboratory (NLR).

Pijpers, E.W. & Van Zutphen, W.J.C.M. (1995). *Development and simulator implementation of an automatic speech recognition application for the mid-life update F-16 Cockpit* (Report NLR CR 95220 L **II.3**). Amsterdam: National Aerospace Laboratory (NLR).

Prévôt, T. & Onken, R. (1995). In-flight evaluation of CASSY: A system providing intelligent on-board pilot assistance. *Air Traffic Control Quarterly, 3 (3),* 183-204.

South, A. (1996). Effect of Aircraft Performance on Voice Recognition systems. *Proc. AGARD Symposium "Audio effectiveness in aviation"*, Copenhagen (in press).

Steeneken, H.J.M. (1995). *Development and performance of a cockpit control system operated by voice; summary report on phase II (DMKLu/AC02/A/9105-Report II.0)* (Report TNO-TM 1995 A-42 **II.0**). Soesterberg: TNO Human Factors Research Institute.

Steeneken, H.J.M. (1996). *Development and performance of a cockpit control system operated by voice; summary report of project DMKLu/AC02-A/9105* (Report TM-96-A055 **III.0**). Soesterberg: TNO Human Factors Research Institute.

Steeneken, H.J.M. & Kriekaard, J.J. (1995a). *Development and evaluation of the electro-acoustical input environment for an automatic speech recognizer in cockpit applications* (Report TNO-TM 1995 A-40 **II.1**). Soesterberg: TNO Human Factors Research Institute.

Steeneken, H.J.M. & Kriekaard, J.J. (1995b). *Prediction of the performance of automatic speech recognition for control applications in a fast-jet cockpit* (Report TNO-TM 1995 A-41 **II.2**). Soesterberg: TNO Human Factors Research Institute.

Steeneken, H.J.M., Kriekaard, J.J. & Van Leeuwen, D.A. (1996a). *Automatic speech recognition performance in a simulation-based fast-jet cockpit application* (Report TM-96-A053 **III.1**). Soesterberg: TNO Human Factors Research Institute.

Steeneken, H.J.M., Kriekaard, J.J. & Van Leeuwen, D.A. (1996b). *Spontaneous-speech data base for cockpit control applications applied to commercial state-of-the-art speech recognition technology* (Report TM-96-A054 **III.2**). Soesterberg: TNO Human Factors Research Institute.

Steeneken, H.J.M. & Pijpers, E.W. (1996). Development and Performance of a Cockpit control system operated by Voice. *Proc. AGARD Symposium "Audio effectiveness in aviation"*, Copenhagen (in press).

Steeneken, H.J.M. & Van Velden, J.G. (1989a). RAMOS—Recognizer assessment by means of manipulation and speech. *European Speech Conf. ESCA,* Paris.

Steeneken, H.J.M. & Van Velden, J.G. (1989b). Objective and diagnostic assessment of (isolated) word recognizers. *Proc. IEEE ICASSP,* Glasgow, pp.540-543.

Steeneken, H.J.M. & Van Velden, J.G. (1991). RAMOS recognizer assessment by means of manipulation of speech applied to connected speech recognition. *Proc. Eurospeech'91,* Genoa, pp.529-532.

Van Velden, J.G. & Houtgast, T. (1992). *Laboratory evaluation of five automatic speech recognition systems* (Report IZF 1992 A-26). Soesterberg: TNO Institute for Perception.

Williamson, D.T. (1996). Flight Test Performance Optimization of ITT VRS-1290 Speech Recognition System. *Proc. AGARD Symposium "Audio effectiveness in aviation",* Copenhagen (in press).

Soesterberg, 29 November 1996

Dr. ing. H.J.M. Steeneken
(author, project manager)

VERZENDLIJST

| | |
|---|---|
| 1. | Directeur M&P DO |
| 2. | Directie Wetenschappelijk Onderzoek en Ontwikkeling Defensie |
| 3. { | Hoofd Wetenschappelijk Onderzoek KL<br>Plv. Hoofd Wetenschappelijk Onderzoek KL |
| 4. | Hoofd Wetenschappelijk Onderzoek KLu |
| 5. { | Hoofd Wetenschappelijk Onderzoek KM<br>Plv. Hoofd Wetenschappelijk Onderzoek KM |
| 6, 7 en 8. | Bibliotheek KMA, Breda |
| 9 t/m 13. | Maj.ir. J.H.J. Verhulst, Koninklijke Luchtmacht, Afdeling Wetenschappelijke Ondersteuning, DM/MXS, Den Haag |

Extra exemplaren van dit rapport kunnen worden aangevraagd door tussenkomst van de HWOs of de DWOO.