NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

AC/323(IST-031)TP/57

www.rto.nato.int

RTO TECHNICAL REPORT **TR-IST-031**

Speech Processing in Realistic Battlefield Environments

(Le traitement de la parole en environnement de combat réaliste)

This Technical Report has been prepared as a result of a project on "Speech Processing Using Realistic Battlefield Data" for the RTO Information Systems Technology Panel (IST) by Task Group 013.

Published April 2009

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
APR 2009	**N/A**	

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Speech Processing in Realistic Battlefield Environments (Le traitement de la parole en environnement de combat réaliste)	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Research and Technology Organisation North Atlantic Treaty Organisation BP 25, F-92201 Neuilly-sur-Seine Cedex, France**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**SAR**	**48**	

NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

AC/323(IST-031)TP/57

www.rto.nato.int

RTO TECHNICAL REPORT **TR-IST-031**

Speech Processing in Realistic Battlefield Environments

(Le traitement de la parole en environnement de combat réaliste)

This Technical Report has been prepared as a result of a project on "Speech Processing Using Realistic Battlefield Data" for the RTO Information Systems Technology Panel (IST) by Task Group 013.

The Research and Technology Organisation (RTO) of NATO

RTO is the single focus in NATO for Defence Research and Technology activities. Its mission is to conduct and promote co-operative research and information exchange. The objective is to support the development and effective use of national defence research and technology and to meet the military needs of the Alliance, to maintain a technological lead, and to provide advice to NATO and national decision makers. The RTO performs its mission with the support of an extensive network of national experts. It also ensures effective co-ordination with other NATO bodies involved in R&T activities.

RTO reports both to the Military Committee of NATO and to the Conference of National Armament Directors. It comprises a Research and Technology Board (RTB) as the highest level of national representation and the Research and Technology Agency (RTA), a dedicated staff with its headquarters in Neuilly, near Paris, France. In order to facilitate contacts with the military users and other NATO activities, a small part of the RTA staff is located in NATO Headquarters in Brussels. The Brussels staff also co-ordinates RTO's co-operation with nations in Middle and Eastern Europe, to which RTO attaches particular importance especially as working together in the field of research is one of the more promising areas of co-operation.

The total spectrum of R&T activities is covered by the following 7 bodies:

- AVT Applied Vehicle Technology Panel
- HFM Human Factors and Medicine Panel
- IST Information Systems Technology Panel
- NMSG NATO Modelling and Simulation Group
- SAS System Analysis and Studies Panel
- SCI Systems Concepts and Integration Panel
- SET Sensors and Electronics Technology Panel

These bodies are made up of national representatives as well as generally recognised 'world class' scientists. They also provide a communication link to military users and other NATO bodies. RTO's scientific and technological work is carried out by Technical Teams, created for specific activities and with a specific duration. Such Technical Teams can organise workshops, symposia, field trials, lecture series and training courses. An important function of these Technical Teams is to ensure the continuity of the expert networks.

RTO builds upon earlier co-operation in defence research and technology as set-up under the Advisory Group for Aerospace Research and Development (AGARD) and the Defence Research Group (DRG). AGARD and the DRG share common roots in that they were both established at the initiative of Dr Theodore von Kármán, a leading aerospace scientist, who early on recognised the importance of scientific support for the Allied Armed Forces. RTO is capitalising on these common roots in order to provide the Alliance and the NATO nations with a strong scientific and technological basis that will guarantee a solid base for the future.

Published April 2009

ISBN 978-92-837-0060-9

Table of Contents

	Page
List of Tables	**v**
Preface	**vi**
Foreword	**vii**
Programme Committee	**viii**
Executive Summary and Synthèse	**ES-1**
Chapter 1 – Introduction	**1-1**
1.1 Military Importance	1-1
1.2 Technical Challenge	1-1
1.3 Work Program	1-1
1.4 Report Organization	1-2
Chapter 2 – Military Speech Databases	**2-1**
2.1 Introduction	2-1
2.2 Terminology	2-1
2.2.1 Language Specification	2-1
2.2.2 Language Proficiency	2-2
2.3 A Selection of Available Military Speech Databases	2-2
2.3.1 FELIN Database	2-2
2.3.1.1 Overview	2-2
2.3.1.2 Technical Specifications	2-3
2.3.1.3 Limitations of Use	2-3
2.3.2 Canadian Soldier System Database	2-3
2.3.2.1 Overview	2-3
2.3.2.2 Technical Specifications	2-4
2.3.2.3 Limitations of Use	2-4
2.3.3 Dismounted Close Combat Database (DCCD)	2-5
2.3.3.1 Overview	2-5
2.3.3.2 Technical Specifications	2-5
2.3.3.3 Limitations of Use	2-6
2.3.4 Non-Native Military Air Traffic Control (nnMATC) Database	2-6
2.3.4.1 Overview	2-6
2.3.4.2 Characteristics	2-6
2.3.4.3 Technical Specifications	2-6
2.3.4.4 Limitations of Use	2-7
2.3.5 Non-Native Civilian Air Traffic Control (nnCATC) Database	2-7
2.3.5.1 Overview	2-7
2.3.5.2 Characteristics	2-7
2.3.5.3 Technical Specifications	2-7
2.3.5.4 Limitations of Use	2-8

2.3.6 Destined Glory 04 Database (DG04DB) 2-8
2.3.6.1 Overview 2-8
2.3.6.2 Technical Specifications 2-9
2.3.6.3 Limitations of Use 2-9
2.3.7 KFOR Text Corpus 2-9
2.3.7.1 Overview 2-9
2.3.7.2 Technical Specifications 2-10
2.3.7.3 Limitations of Use 2-10
2.4 References 2-10

Chapter 3 – Task Definitions and Metrics 3-1
3.1 Introduction 3-1
3.2 Speech-to-Text 3-1
3.3 Call-Sign Identification 3-1
3.4 Entity Clustering Tasks 3-2
3.4.1 Speaker Entity Identification 3-3
3.4.2 Listener Entity Identification 3-3
3.5 Native/Non-Native Detection (N3D) 3-3
3.6 Processing of Data 3-4
3.7 Evaluation of Human Baseline 3-4
3.8 References 3-5

Chapter 4 – Experimental Results 4-1
4.1 Introduction 4-1
4.2 The Impact of Battlefield Speech 4-1
4.2.1 Impact on Speech Recognition 4-1
4.2.2 Impact on Speaker Recognition 4-1
4.2.3 Impact on Entity Recognition 4-2
4.3 Speech Recognition Experiments on the nnMATC Corpus 4-2
4.3.1 Evaluation using HTK and Sphinx Recognizers 4-3
4.3.1.1 Sphinx 4-3
4.3.1.2 SphinxTrain 4-3
4.3.1.3 Force Alignment 4-4
4.3.1.4 Sphinx 3 Decoder 4-4
4.3.1.5 Language Models 4-4
4.3.1.6 Results Obtained 4-4
4.3.1.7 HTK 4-5
4.3.2 Evaluation using DGA Recognizer 4-5
4.3.2.1 Experiment Descriptions 4-5
4.3.2.2 Results Summary 4-6
4.3.2.3 Improvements 4-7
4.3.3 Evaluation using VOGON Recognizer 4-7
4.4 Conclusions 4-8
4.5 References 4-8

Chapter 5 – Recommendations and Conclusions 5-1
5.1 References 5-1

List of Tables

Table		Page
Table 1	Lexical Transcription Errors	4-2
Table 2	ATC Data Partitions	4-3
Table 3	DGA Word Error Rates on Test Partitions	4-7
Table 4	VOGON Word Error Rate on Test Partitions	4-7

Preface

Communications, command and control, intelligence, and training systems are more and more making use of speech technology components: i.e., speech coders, voice controlled C^2 systems, speaker and language recognition, and automated training suites. Interoperability of these systems is not a simple standardization problem as the speech of each individual user is an uncontrolled variable such as non-native speakers using, in addition to their own language, an official NATO language. For multi-national military operations, this may cause a reduced performance or even cause malfunction of an action. Standardized assessment methods and specifications both for commercial-off-the-shelf (COTS) and for development of new technology are required. The work was separated into four tasks:

1) Collect native and non-native unclassified and classified speech communications from training exercises and actual operations;
2) Produce annotated database(s) that might be used beyond the confines of the Task Group;
3) Assess effects on performance of recognisers and communication equipment; and
4) Relate derived results to military applications.

In this report the results of the study are presented.

Foreword

Efficient speech communication is recognized as a critical and instrumental capability in many military applications such as command and control, aircraft and vehicle operations, military communication, translation, intelligence, and training. The former NATO research study group on speech processing (AC243 (Panel 3) RSG10) conducted since its establishment in 1978 experiments and surveys focused on military applications of language processing. Guided by its mandate, the former RSG10 initiated in the past the publication of overviews on potential applications of speech technology for military use and also organized several workshops and lecture series on military-relevant speech technology topics. Recently the group continued under the IST Panel as AC232/IST/TG001.

In recent years, the speech R&D community has developed or enhanced many technologies which can now be integrated into a wide-range of military applications and systems:

- Speech coding algorithms are used in very low bit-rate military voice communication systems. These state-of-the-art coding systems increase the resistance against jamming;
- Speech input and output systems can be used in control and command environments to substantially reduce the workload of operators. In many situations operators have busy eyes and hands, and must use other media such as speech to control functions and receive feedback messages;
- Large vocabulary speech recognition and speech understanding systems are useful as training aid and to prepare for missions;
- Speech processing techniques are available to identify talkers, languages, and keywords and can be integrated into military intelligence systems; and
- Automatic training systems combining automatic speech recognition and synthesis technologies can be utilized to train personnel with minimum or no instructor participation (e.g., Air traffic controllers).

This report is the result of a project on “Speech Processing Using Realistic Battlefield Data” with contributions of all Task Group members, which represent nine NATO countries (Belgium, Canada, France, Germany, Netherlands, Spain, Turkey, the United Kingdom, and the United States). Because speech technologies are constantly improving and adapting to new requirements, it is the intention of the Task Group to initiate projects on military applications of speech technology. Therefore the group appreciates any comment and feedback on this report.

Programme Committee

Membership of Information Systems Technology Research Task Group 013

CHAIRMAN

Dr. Timothy ANDERSON
Air Force Research Laboratory, AFRL/HECA
2255 H Street
Wright Patterson AFB
OH 45433-7022
USA

SECRETARY

Mr. Carl SWAIL
National Research Council
Flight Research Laboratory
Building U-61, Montreal Road
Ottawa, Ontario K1A 0R6
CANADA

MEMBERS

BELGIUM

Dr. Stéphane PIGEON
Koninklijke Militaire School
Leerstoel voor Telecommunicaties
Royal Military Academy
Renaissancelaan 30
B-1000 Brussels

FRANCE

Mr. Mathieu MANTA
DGA/DET/CEP
16 bis avenue Prieur de la Côte d'Or
94114 Arcueil Cedex

GERMANY

Dr. Matthias HECKING
FGAN/FKIE
Neuenahrer Str 20
D-53343 Wachtberg-Werthhoven

NETHERLANDS

Dr. David A. van LEEUWEN
TNO Human Factors
P.O. Box 23
3769 ZG Soesterberg

SPAIN

Dr. Juan GOMEZ-MENA
CIDA-SDGTECEN-DGAM
c/Arturo Soria, 289
Madrid – 28033

TURKEY

Mr. Mehmet Ugur DOGAN
TUBITAK-UEKAE
National Research Institute of Electronics & Cryptology
P.K. 74
41470 Gebze. Kocaeli

UNITED KINGDOM

Mr. Paul COLLINS
Room G007, A2 Building
DSTL Farnborough
Hampshire GU14 0LX

UNITED STATES

Mr. John J. GRIECO
AFRL/IFEC
525 Brooks Rd.
Rome, NY 13441

Dr. Aaron D. LAWSON
RADC
525 Brooks Rd.
Rome, NY 13441

Dr. Wade SHEN
Information Systems Technology Group
MIT Lincoln Laboratory
244 Wood Street
Lexington, MA 02420-9108

Speech Processing in Realistic Battlefield Environments

(RTO-TR-IST-031)

Executive Summary

Multilingual speech and language technology is becoming recognized as an important issue for international organizations, both civilian and military. For instance, one might want to use a speech coder optimized for French in Germany or Turkey. A native speaker of Spanish might want to use a speech recognizer trained for American English in a military noise environment. Additionally with the explosion of multilingual text material on the web, a British user might want to access Dutch documents using English search terms. For reasons such as these, a special task group of the NATO Research and Technology Organisation (RTO) started a project on the development and assessment of multilingual speech and language applications.

To stimulate research and evaluation the NATO Research Task Group on Speech and Language Technology (IST-031/TG-013) collected databases of speech under battlefield conditions. These databases are the Non-Native Civilian Air Traffic Control Speech Corpus, Destined Glory NATO Military Exercise Corpus, and the Non-Native Military Air Traffic Control Speech Corpus. Studies conducted by participating NATO laboratories and discussed here suggest that many Commercial-off-the-shelf (COTS) speech systems, which were designed for civilian conditions cannot be effectively used for battlefield conditions. The main findings and recommendations are:

- It is suggested that the effect of non-native speech in battlefield environments on the speech production quality is likely to be detrimental to the effectiveness of communication in general, in particular to the performance of communication equipment and weapon systems equipped with vocal interfaces (e.g., advanced cockpits, command, control, and communication systems, and information warfare).
- Commercial-off-the-shelf speech recognition systems are not yet able to address the wide speaker variability associated with non-native speech in battlefield environments.
- Databases obtained or collected during this study have been distributed to all participating NATO countries, and most are available in CD-ROM format to interested parties.
- Progress in the field of military based speech technology, including advances in speech based system design has been restricted due to the lack of availability of databases of non-native speech in battlefield environments.
- It is foreseen that in the future it will be necessary to improve the coordination of multi-national military forces. The need therefore exists for planned simulations of military personnel using a wide range of speech technology.
- Military operations are often conducted under conditions of stress induced by high workload, sleep deprivation, fear and emotion, confusion due to conflicting information, psychological tension, pain, and other typical conditions encountered in the modern battlefield context. These conditions on top of effects on non-native speech in battlefield environments are sure to challenge speech technology into the future.

Le traitement de la parole en environnement de combat réaliste

(RTO-TR-IST-031)

Synthèse

La technologie de la parole et du langage multilingue commence à être reconnue comme un enjeu important par les organisations internationales civiles et militaires. Par exemple, on peut demander à utiliser un codeur vocal optimisé pour le Français en Allemagne ou en Turquie. Un hispanophone peut demander à utiliser un dispositif de reconnaissance vocale qualifié pour de l'anglo-américain dans un environnement militaire bruyant. De plus, avec l'explosion des textes multilingues sur internet, un utilisateur Britannique peut demander d'accéder à des documents Hollandais en utilisant des termes de recherche Anglais. Pour de telles raisons, un groupe opérationnel spécial de l'Organisation de la Recherche et de la Technologie (RTO) de l'OTAN a débuté un projet sur le développement et l'évaluation des applications multilingues de la parole et du langage.

Pour stimuler la recherche et l'évaluation, le Groupe Opérationnel de Recherche sur la Parole et le Langage (IST-031/TG-013) de l'OTAN a collationné des données vocales en environnement de combat. Ces bases de données sont le *Non-Native Civilian Air Traffic Control Speech Corpus,* le *Destined Glory NATO Military Exercise Corpus,* et le *Non-Native Military Air Traffic Control Speech Corpus*. Des études conduites par des laboratoires associés à l'OTAN et discutées ici suggèrent que de nombreux systèmes vocaux disponibles dans le commerce (COTS), conçus pour les civils en temps de paix, ne peuvent pas être efficacement utilisés au combat. Les principales conclusions et recommandations sont les suivantes :

- On pense que l'usage d'une langue non-maternelle dans un environnement de champ de bataille a un effet sur la qualité de la production vocale et peut être dommageable à l'efficacité de la communication en général, et en particulier pour les performances des équipements de communication et des systèmes d'armes équipés d'interfaces vocales (par exemple les cockpits évolués, les systèmes de commandement, de contrôle et de communication et de guerre de l'information).
- Les systèmes de reconnaissance du commerce ne sont pas encore capables de prendre en compte les différents modes d'utilisation dus à l'usage d'une langue non maternelle en environnement de combat.
- Les bases de données obtenues ou recueillies durant cette étude ont été distribuées à toutes les Nations participantes de l'OTAN et sont disponibles pour la plupart sur format CD-ROM pour les intéressés.
- Les progrès dans le domaine de la technologie vocale militaire, incluant des avancées dans la conception des systèmes vocaux, ont été limités à cause du manque de disponibilité des bases de données d'expression en langue non maternelle en environnement de combat.
- Il est à prévoir que dans le futur, il sera nécessaire d'améliorer la coordination des forces militaires multinationales. En conséquence, le besoin se fait sentir de simulations programmées de personnels militaires utilisant une grande variété de technologies de langages.
- Les opérations militaires sont souvent conduites dans des conditions de stress induites par la forte charge de travail, la privation de sommeil, la peur et l'émotion, la confusion due aux informations

contradictoires, la tension psychologique, la souffrance et autres conditions rencontrées sur le champ de bataille moderne. Ces conditions, qui sont parmi les premières à influer sur l'expression dans une langue non maternelle au combat seront assurément un enjeu de la technologie du langage dans le futur.

Chapter 1 – INTRODUCTION

1.1 MILITARY IMPORTANCE

As speech-processing technology becomes mature, the potential to utilize the technology for speech-enabled military systems strongly increases. The technology can be embedded in military communication, command and control, intelligence, and training systems. Interoperability of these systems is paramount to the success of NATO multi-national operations. This, however, creates interesting and unique problems in the successful implementation of speech technology, where multi-national forces working in a coalition environment exist. In this environment, speech-processing equipment designed by one country must be used by soldiers from another. Unlike other military systems, where interoperability could be created by simply rewriting a user's manual in the native language for a particular soldier, speech systems must be created and measured for effectiveness before deployment. Interoperability of military systems such as speech coders, voice controlled C2 systems, speaker and language recognition, and automatic training suites are not a simple standardization problem. The speech of each individual user is an uncontrolled variable. The use of speech systems by non-native speakers speaking the official NATO languages, French and English, may cause reduced performance or even complete malfunction of a system**,** especially in a battlefield environment. Standardized assessment methods, specifications, and training techniques are required for both commercial-off-the-shelf (COTS) and for the development of new technology-based military systems.

1.2 TECHNICAL CHALLENGE

The IST-RTG013 recognized the need to perform research and studies on this topic to better understand, detect, and mitigate the effects of native and non-native speech production in military battlefield environments. Minimal research had been conducted in this area prior to the initiation of this project. Commercial systems were built with little regard for non-native speech production and battlefield noise and channel conditions. As a result, interoperability of systems developed for specific languages becomes an issue, especially when military forces are pressed into action often with short notice. Examine the case where in a particular operation a native speaker of Dutch speaking Dutch must use a speech coder in a secure communication device, which was optimized for British English. Imagine the case where a native speaker of German might need to use a speech translator trained for Spanish. Interoperability of speech systems is an important issue for many applications of modern speech technology in the coalition environment. For this reason, the NATO Research and Technology Organization (RTO) under the Information Systems Technology (IST) Panel authorized a task group to identify the application of and assess the use of multilingual speech technology in the military battlefield environment.

1.3 WORK PROGRAM

In the past, TG01 constructed projects which studied the various effects of military environments in relation to the performance of speech technology. Examples are the effect of noise on speech recognition, the effect of stress induced by workload, sleep deprivation, and battlefield stress. The biggest impact of these projects was the creation of datasets representative of the military environment, which fostered interest in the academic and industrial scientific communities. This has shaped the development and evaluation of speech technology for the harsh military environment.

Speech data was collected in three conditions representative of military battlefield conditions to foster research on multi-lingual, non-native speech in battlefield conditions. This data set is very representative of military type

communication in a military battlefield scenario, and was used for evaluation and modification of Automatic Speaker Recognition and Word Spotting. These databases also focused research on non-native speech and robustness issues, which led to a special session at the international speech and language conference, Interspeech 2007 in Antwerp, Belgium.

1.4 REPORT ORGANIZATION

This report is organized into five chapters. Below is a description of the content in each chapter.

Chapter 1:

This chapter contains an introduction to the project and describes how the report is organized.

Chapter 2:

This chapter presents the various military databases which were considered and/or collected for this project. Also included in this chapter is a detailed description of these databases. An overall description of each database and its content, amount of data, language, non-native type, and characteristics is included.

Chapter 3:

This chapter presents an experimental plan on using the database (nnMATC) to measure the performance of speech processing systems on the problems of detection, classification, and assessment of non-native, accented speech in a realistic battlefield environment.

Chapter 4:

The issues and findings of various speech systems are presented.

Chapter 5:

In this chapter conclusions are drawn. A discussion of the impact that multi-lingual and non-native speech in a realistic battlefield environment has on military speech technology and its application is presented.

Chapter 2 – MILITARY SPEECH DATABASES

2.1 INTRODUCTION

Over the years, pronunciation variation due to non-native speech has been the interest of many phoneticians. A most remarkable number of researchers have studied the production and perception of the famous 'l-r confusion' for Chinese and Japanese natives. It is not an exaggeration to claim[1] that more than 50% of the research papers on non-native speech deal with this interesting subject of /l/ and /r/.

Long after phoneticians were drawn by the subject, non-native speech slowly started to become an issue in speech technology. Since speech databases are an invaluable resource for researchers in the field of speech technology, soon the first non-native speech databases were recorded. One of the problems with non-native speech is that there are potentially so many different kinds: if N is the number of languages in the world[2], the number of non-native accents is close to N^2. This number only considers speech production. If the perception of speech is taken into account too, the number of possible language combinations scale with N^3.

It is clear that in a field of research that has only recently started and with so many possible language combinations, the coverage by speech databases is rather limited. Yet, there are a number of interesting recordings available.

2.2 TERMINOLOGY

2.2.1 Language Specification

In non-native speech communication there are at least three languages of importance:

1) The native language of the speaker (*S*);
2) The language that is spoken, or the target language (*T*); and
3) The native language of the listener (*L*).

In literature one often finds the symbols L1 for native language and L2 for the target language, but this is not always used consistently and, especially in listening experiments, this terminology might lead to confusion. It is always a good exercise to really understand the configuration of languages in a description of non-native language experiments in a research paper. Van Wijngaarden [1] proposes the notation for communication between two persons:

$$S > (T) > L$$

Meaning that a speaker who's native language is *S* speaks in language *T* to a listener whose native language is *L*. For the purpose of speech databases for speech technology, it usually suffices to specify *S* and *T* only, because *L* is either not considered or the technology takes the role of the listener, as is the case for speech, speaker, language and accent recognition.

[1] Based on JASA abstracts.

[2] *N* is estimated to be about 6000; the bible has been translated into 2000 languages alone.

2.2.2 Language Proficiency

One of the most important parameters in non-native speech communication is the language proficiency of the speaker and listener. There is a NATO standard, STANAG 6001, that classifies the language proficiency of people into five levels:

1) Elementary;
2) Fair (Limited working);
3) Good (Minimum professional);
4) Very good (Full professional); and
5) Excellent (Native/Bilingual).

These levels define both speaking and listening proficiency.

For speech databases, it is very important that the language proficiency of the individual speakers be known, because the quality and character of the speech is very dependent on this. None of the databases discussed in this chapter has classified the speakers according to STANAG 6001 levels, but there is generally information about the speaker's non-native language acquisition. Important information includes:

- **Native language:** The mother tongue of the speaker;
- **Age of acquisition:** The speaker's age when the non-native language was learned; and
- **Experience:** The number of years that the speaker has been regularly using the language.

Generally, an age of acquisition of over 6 years is considered to always lead to a noticeable non-native accent. All things being equal, the higher the age of first learning, the stronger the non-native accent will be. Of course, these parameters are not the only important factors in language proficiency; there are also matters such as willingness to learn another language, level of exposure, talent, etc. There are numerous cases where 'expatriates' live in a foreign country for decades without being exposed to the local native language at all.

Databases differ in the information that is specified about the speakers, even though there has been an effort to guide the database meta-data collection, such as through the EAGLES handbook [2].

2.3 A SELECTION OF AVAILABLE MILITARY SPEECH DATABASES

In this section an overview is given of some military databases that are available. There are many more recordings made of utterances under a wide range of conditions, for all the different studies made in literature. Here we only list the databases that have some relevance to speech technology research in military battlefield conditions.

2.3.1 FELIN Database

2.3.1.1 Overview

FELIN is a French database that was recorded in order to evaluate:

- How French infantrymen use speech-based command and control systems; and
- The performance of commercial systems on such a task.

The recording campaign lasted 3 days, involved 6 infantrymen in 11 exercises. The scenario was urban reconnaissance.

Each infantryman had his own recording device. Each device had 2 microphones: a bone-conducting microphone located at the top of the head inside the helmet and a boom-mounted microphone fixed on the helmet close to the cheek. One of the devices used was a speech command system. It recognized commands and replied using speech synthesis when needed. For the other devices, the speech recognition was done *a posteriori*.

All speech commands were preceded by the keyword "vocal" to initiate a dialog with the speech command system. The speech commands are divided in two types: those available for the whole group and those available for the group leader.

2.3.1.2 Technical Specifications

The database is made up of 15 hours 20 minutes of audio recordings.

All recordings have been transcribed with the annotation tool, Transcriber[3], and saved in TRS format. All non-speech events are annotated.

Audio format: WAV – 16 kHz – 16 bits

Channel 0: bone-conducting microphone

Channel 1: boom-mounted microphone

2.3.1.3 Limitations of Use

The database distribution and usage is provided to the following restrictions:

- The database can only be used for academic, research and evaluation purposes;
- The database or the results of its use, cannot be used for any commercial purpose;
- The database cannot be redistributed without authorization of the database owner[4];
- Any publication or evaluation result concerning the database must be communicated to the database owner;
- The anonymity of the recorded people must be guaranteed in any publication relative to the database; and
- The database must be destroyed upon request of its owner.

2.3.2 Canadian Soldier System Database

2.3.2.1 Overview

Experiments were carried out at Defence Research and Development Canada (DRDC) looking at communications among ground forces in an urban warfare environment as part of a future soldier system.

[3] http://trans.sourceforge.net/.

[4] Owner = DGA.

Audio recordings were made of the communications during these exercises. The exercises were done at two locations: in the laboratory using a computer gaming simulation and live exercises at a military facility.

The data was segmented into files with individual transmissions. All the files are labelled by time and speaker identification. The speaker id can be a number or a role, for example “2IC” for second in command. The data is divided into two directories for the laboratory simulation exercises and two directories for the field exercises. Each directory has up to 65 session directories with a number of individual transmission files in each. Some of the session directories contain log files with information about each transmission, but since the same information is contained in the file name there is no new information in the file.

The speakers are predominantly males; however, there are a few females in the dataset. There are 35 unique speakers in the field sessions and 7 speakers in the laboratory sessions. There is probably an overlap between the two groups, but this is unknown.

The laboratory data was recorded using head-set mounted microphones. Collection of the field data used a head-set microphone attached to a personal communications set. The recordings were done at a base station.

2.3.2.2 Technical Specifications

The data files are in PC WAV format with the following specifications:

- Encoding: linear PCM
- Sample rate: 44.1 kHz
- Resolution: 16 bits / sample
- Channels: 1

In order to maintain the time sequence of the files and to indicate the speaker, this information was included in the filename. The following filename format was used for the .WAV files for the laboratory sessions:

END TIME + PLAYER NAME + <space> + CHANNEL # + “.WAV”

END TIME = 6 digit time code when transmission ended, hhmmss

PLAYER NAME = “SC”, “2IC”, or “Player 2” to “Player 7”

CHANNEL # = 1 digit code

The filename format for the field experiments is similar and includes:

END TIME = 6 digit time code when transmission ended, hhhmmmsss

PLAYER NAME = P##G# where G is the group number and P is the person number.

2.3.2.3 Limitations of Use

There are no restrictions on the distribution of this database; however, the database can only be used for research purposes. The producers of the data have asked that the National Research Council, Institute for Aerospace Research co-ordinate its distribution.

2.3.3 Dismounted Close Combat Database (DCCD)

2.3.3.1 Overview

The DCCD database was recorded between July and September 2001 by Aurix Ltd. (then called 20/20 Speech Ltd) for an UK MoD project for research into speech interfaces for dismounted infantry.

The trial data contains 16 male speakers who were soldiers in the British Army at the time of the recordings. The recordings consist of read speech (UK-English) consisting of alpha-numeric strings (using UK generic "number plate" format) and command-and-control type phrases. The speech was recorded during periods of external noise (e.g., gunfire) and of physical stress and (e.g., jogging). There are 11 distinct "acoustic" regions:

01) No additional noise; no physical stress.
02) Reverberation (Butts); no physical stress.
03) Gunfire and reverberation (Butts); no physical stress.
04) Gunfire (Firing-line); no physical stress.
05) Interfering speech; no physical stress.
06) Engine noise; no physical stress.
07) No additional noise; crawling and speaking.
08) No additional noise; jogging and speaking.
09) No additional noise; recovering and speaking.
10) No additional noise; speaking quietly.
11) No additional noise; shouting.

2.3.3.2 Technical Specifications

There are 47 speech files (dcc016-dcc063). Each speech file was recorded from a UK Infantry standard microphone as used for the personal role radio (this consists of a good quality noise-cancelling boom microphone that can be adjusted close to the speaker's lips). The recordings were made on a DAT recorder at 44.1 kHz and later down-sampled to 20.50 kHz 16 bits PCM single channel.

The material spoken consisted of two types:

1) **Sighting Reports** consisting of reading between 1 and 6 "targets" each represented by the form of a string *"alpha one two three bravo charlie delta"* with report start and end notation. The following example is for 2 targets:

   "target report begins, two targets sighted"

   "charlie nine seven three juliet oscar bravo"

   "juliet nine seven one sierra quebec uniform"

   "report ends"

2) **Control Words** consisting of 30 phrases, each phrase consisting of between one and three words. Some examples are:

   "left" "image" toggle" "select" "mode" "transmit image" "toggle map image"

The speech files are currently recorded in a raw format and transcribed with the standard Transcriber tool. They occupy about 10 CDs. There is around 5.5 hours of speech comprising around 39,000 words in 10,575 phrases.

2.3.3.3 Limitations of Use

The database has not been used or released outside of Aurix Ltd. If this was required, it would need some additional work to formalize the documentation and double-check the directory structures.

2.3.4 Non-Native Military Air Traffic Control (nnMATC) Database

2.3.4.1 Overview

The Non-Native Military Air Traffic Communications (nnMATC) database was collected in Spring 2005 in the framework of the NATO/IST-013/RTG-031 task group, to support ongoing research in the field of speech processing under realistic battlefield conditions. Among those conditions, speakers non-nativeness and channel noise most heavily affect speech recognition performance. The nnMATC database combines the adverse effects of non-native speech and noisy environment, through realistic air traffic control communications recorded from an operational military Air Traffic Control (ATC) center. The nnMATC has been recorded at the time when it wasn't clear yet if we would succeed in releasing the Civilian ATC database (nnCATC) to the group. The nnMATC offers various advantages over the nnCATC, in particular it is a real military environment.

2.3.4.2 Characteristics

The nnMATC database consists of 24+ hours of contiguous ATC communications. These recordings were taped from the ATC center, implying a different speech quality depending on the speaker's location: on the controller-side, speech is mostly clean; on the pilot side, recordings suffer from a combination of background noise (cockpit) and communication interferences.

The non-native English accents covered on the controller side are mainly Belgian Dutch and Belgian French. On the pilot side, the variety is much wider with – among others – Dutch, Belgian Dutch, French, Belgian French, German, Italian, and Spanish accents. A few native American, British and Canadian English speakers are represented among the pilots as well. Most speakers are males, although a few female speakers are present as well, mainly among the controllers.

The nnMATC database was acquired through 13 sessions, all taking place at the same location but on different days. Multi-channel sessions – there are 12 of them – consist of multiple channels that have been recorded simultaneously for a few consecutive hours (typically 3-5 hours). Depending on air traffic conditions, 5 to 9 of those channels were active. The 13th session differs from the others as it relates to the recording of a single channel across a much longer period of time (a few days).

Most of these recordings originally suffered from long periods of inactivity – often up to several minutes. Therefore, silences have been trimmed down to 2 seconds. This process ended up in squeezing approximately 700 hours of raw audio material into 24+ hours of gapless speech (total time across all sessions and channels).

2.3.4.3 Technical Specifications

Audio bandwidth: 300 Hz – 3400 Hz (tapped over phone lines)

Recoding format:	wav, 22.05 kHz, 16-bit linear
Silence trimming:	Signal < –40dbFS with a 1s post- and pre-roll margin
Total recording time:	24:34:04 (hh:mm:ss)
File format:	nnMATC_sessionID_frequency.WAV (multi-session)
	nnMATC_frequency_fileID.WAV (single session)

2.3.4.4 Limitations of Use

The nnMATC database is NATO UNCLASSIFIED. However, limitations of use apply. The nnMATC database is primary intended for research purposes within the NATO/IST-013/RTG-031 Group. Commercial use of the database is strictly prohibited. Identities involved in the recordings should be kept anonymous in any unclassified publication.

2.3.5 Non-Native Civilian Air Traffic Control (nnCATC) Database

2.3.5.1 Overview

The Non-Native Civilian Air Traffic Communications (nnCATC) was meant to be the reference ATC database for the NATO/IST-013/RTG-031 task group. As civilian data was supposed to be less sensitive to usage restrictions than its military counterpart, there was no plan to collect our own (military) database at the beginning of the project, but rather to collaborate with a European civilian air traffic control center to get access to ATC communications. Unfortunately, legal issues were more sensitive than expected and after one year of negotiations, no confidence was given that the data could ever be released. At that time, the group started to record its own military database (please refer to the nnMATC, described earlier). Eventually, an agreement was made and the civilian ATC database was released to the group under severe restrictions of use, coincidently with the nnMATC.

Because of those heavy restrictions, we discourage the use of the nnCATC and prefer the nnMATC. The nnCATC remains however available for those interested in additional civilian ATC data.

Compared to the nnMATC, the nnCATC database offers a wider variety in non-native accents and preserves the absolute timeline (in other words, pauses between communications have not been removed). The amount of actual speech data is however much smaller, approximately 8 hours versus 24h.

2.3.5.2 Characteristics

The nnCATC was recorded on 3 radio frequencies (delivery, ground and departure) which are supposed to intercept the same flights. Each frequency was recorded for 20 hours continuously, ending up with 60 hours in total split into 30-minute segments. Each segment holds approximately 4 minutes of speech, on average.

Sound files were transferred from a digital archival system and suffer from audible aliasing (8 kHz – 16 bit). The full database is 3.22 GBytes in size.

2.3.5.3 Technical Specifications

Audio bandwidth:	300 Hz – 3400 Hz (phone line quality)

Recoding format:	wav, 8 kHz, 16-bit
Silence trimming:	None
Total recording time:	60:00 (hh:mm) – no silence trimming
File format:	chXX_msgYY.WAV with XX being the channel number (12, 20 or 46) and YY the message number (1 to 40)

2.3.5.4 Limitations of Use

Limitations of use apply, among others:

- Use of the database shall be restricted to research applications in the field of Speech and Language processing.
- Commercial use of the database shall strictly be prohibited.
- User commits itself to treat the information found in this database with confidentiality: the anonymity of the ATC data provider, the air traffic controllers, the pilots and the airline companies involved in the database shall at all times be guaranteed when disseminating the results of the scientific studies carried on the database.
- User shall not further distribute any of the contents of this database to any third party, without the prior written consent of Royal Military Academy (Belgium), the distributor of this database.
- User shall not publicly publish any part of the communications transcriptions in any form.
- User shall not use the contents of this database to take action against the RMA and/or the ATC Data Provider.

2.3.6 Destined Glory 04 Database (DG04DB)

2.3.6.1 Overview

Destined Glory 2004 (DG04) was a maritime expeditionary exercise conducted by STRIKEFORCESOUTH and includes live fire and a NATO Reaction Force Initial Operational Capability demonstration. The exercise was conducted on Capo Teulada range, an Italian Army armor training area, located on Sardinia's southern tip. The area is extremely remote. The exercise was conducted from 20 September to 16 October in 2004. Nearly 9,500 personnel, 50 ships and 46 aircraft participated. The Maritime and Amphibious forces were from 11 NATO nations.

The database created as a result of participation in the exercise has provided an excellent opportunity to assess next generation speech technology. During the exercise over 100 hours of raw audio were recorded. The exercise consisted of sea, land and air units that were available throughout the exercise. As this was an actual military exercise the primary communication devices were military push-to-talk radios. This imparts a real-world character to the communications, namely:

- Purposeful military verbal exchanges;
- Communication fading and multi-path;
- Communications contaminated with noise and interference;

- One-side of the communication; and
- Military vehicle artifacts.

Along with the environmental effects the NATO exercise allowed for a unique opportunity to capture specific human effects. These effects included personnel under physical stress, non-native speakers of English, and mid-communication language/word switching.

Although English and French are the official languages of NATO many of the communications were in the native language of the exercise participants. As such, not only are there examples of non-native English, but also a fair amount of Spanish, Italian, and Greek.

2.3.6.2 Technical Specifications

The data was accessed utilizing 4 Watkins Johnson (WJ-8611) receivers connected to a high quality 24 channel Mark Of The Unicorn 24I/O analog-to-digital converter connected to PCI-424. All audio files were recorded with 16 bits of fidelity at 48kHz sampling rate in little-endian PCM format. A portion of the data has been transmission marked, with speaker/call-sign markings, and transcribed. This has been accomplished with the Transcriber tool set. To date approximately 2 hours of raw audio has been processed. This contains 53 minutes of packed native/non-native English. Within the database are 57 speakers in 1176 transmissions. These transmissions contain 8277 total words of which 1042 are unique.

The audio file format can be decoded in the following way:

T01_285_AD_0822.pcm

T01 – Identifies the file as a DG04 audio file

285 – Julian date

AD – Air activity frequency D (G indicates ground activity)

0822 – Time of day collection began

The transcription files corresponding to the above audio file are the following format:

T01_285_AD_0822.trs

2.3.6.3 Limitations of Use

The DG04DB is NATO RESTRICTED and as such can only be used for research purposes within the NATO/ IST-013/RTG-031 Task Group. Use outside of the NATO speech research community is strictly prohibited.

2.3.7 KFOR Text Corpus

2.3.7.1 Overview

In the ZENON project [3] an information extraction approach is used for the (partial) content analysis of written English HUMINT reports from the KFOR (Kosovo Force) deployment of the German Federal Armed Forces. Starting point of this development were 4,498 military reports (mostly in English) from the deployment. From these reports 800 were manually annotated and form the *KFOR Corpus* [4]. This corpus is a specialized micro text corpus.

The KFOR corpus is used for the following purposes:

1) It represents the basis for the construction of the information extraction component of the ZENON prototype. The lexicon and the grammars are optimized towards the corpus.
2) The performance of the ZENON information extraction is quantitatively evaluated relative to the KFOR corpus.
3) The KFOR corpus can be used for other research objectives (e.g., complexity of nominal phrases, word sense disambiguation, machine learning of grammatical structures, etc.).

2.3.7.2 Technical Specifications

The used annotation tool is GATE (General Architecture for Text Engineering, http://gate.ac.uk). The corpus covers 886,000 tokens and contains the annotations in different layers. The following layers are available:

- Original markups: In this layer those parts of the message are annotated, which are already formatted (e.g., addressee, topic, source).
- Token: This layer contains the annotations, which are supplied by the Tokenizer and the Part-of-Speech Tagger.
- Gazetteer: In this layer those expressions are annotated, which were identified over lists of names (e.g., first names, city names).
- Sentence: These annotations refer to sentences and begin and end markers of comments.
- Named entities: This layer contains the following annotation types: City, Company, Coordinates, Country, CountryAdj, Currency, Date, GeneralOrg, MilitaryOrg, Number, Percent, Person, PoliticalOrg, Province, Region, River, Time and Title.
- Verb Group: The verbal phrases are annotated.

The corpus is represented in:

- The GATE-specific serial format (‘SerialDataStore’);
- The GATE-specific XML format (‘XML serialization format’);
- The XCES stand-off annotation format; and
- The TIGER-XML format.

2.3.7.3 Limitations of Use

The KFOR Text Corpus is classified (VS-NfD) and is not freely available**.**

2.4 REFERENCES

[1] Wijngaarden, S., Steeneken, H. and Houtgast, T. (2001). “Methods and Models for Quantitative Assessment of Speech Intelligibility in Cross-Language Communication”, In Proc. of the Workshop on Multilingual Speech and Language Processing, Aalborg, Denmark. (Available as NATO Publication RTO-MP-066, April 2003).

[2] Howell, P. (1997). Handbook of Standards and Resources for Spoken Language Systems, Chapter 9: "Assessment methodologies and experimental design", pp. 344-380, Mouton de Gruyter.

[3] Hecking, M. (2006a). "Content Analysis of HUMINT Reports". In: Proc. of the 2006 Command and Control Research and Technology Symposium (CCRTS) "The State of the Art and the State of the Practice", June 20-22, 2006, San Diego, California.

[4] Hecking, M. (2006b). "Das KFOR-Korpus als Ergebnis semantisch annotierter militärischer Meldungen". Forschungsgesellschaft für Angewandte Naturwissenschaft e.V. (FGAN), Wachtberg, Germany, FKIE-Bericht Nr. 124.

Chapter 3 – TASK DEFINITIONS AND METRICS

3.1 INTRODUCTION

The goal of this chapter is to define the evaluation tasks, performance measures, and test corpora to support the evaluation of speech and language processing systems on the NATO IST031/RTG013 Non-Native Military Air Traffic Control (nnMATC) Corpus. Five core speech processing tasks have been defined for evaluating speech and language processing on the nnMATC corpus. While these tasks have been defined specifically with the nnMATC corpus in mind, it is hoped that they might also be applied to other similar corpora. In this section, the task definition and metrics are introduced. Details of the system output formats and the evaluation of system performance are provided in later sections.

For each of the five tasks, a system is provided with two inputs. The first input is a digitally sampled speech waveform file containing multiple speech transmissions. The second input is a human-produced reference segmentation file indicating the starting time and duration of each individual speech transmission. While we realize that most applications of interest would not have the benefit of a reference segmentation file, we believe that systems exploiting readily available side information (e.g., radio frequency energy measurements) could perform high-accuracy automatic segmentation. Giving systems access to the reference segmentation allows us to focus the evaluation on the most interesting speech and language tasks and simplifies our performance metrics.

Though some applications might require systems that process transmissions causally, in this evaluation, we allow systems to process transmissions within a speech file non-causally. A system may both "look ahead" to future transmissions and/or use history information while processing a current transmission.

In all cases, performance is computed by comparing system output to a human-produced reference annotation containing "ground-truth" information.

3.2 SPEECH-TO-TEXT

The goal of the speech-to-text (STT) task is to produce a word transcription of each speech transmission. The system outputs the sequence of words spoken. Non-lexical speaker sounds (e.g., cough, sneeze, breath, lip smack, and laugh) and non-speech sounds (e.g., door slams, tones, etc.) may occur in the speech transmissions, but these sounds should not be scored for evaluation.

System performance will be computed by aligning the sequence of hypothesized words against the sequence of reference words using the alignment strategy found in `sclite` [1]. The word error rate (WER) is the sum of the deletion, insertion and substitution errors divided by the number of words in the reference annotation.

In addition to the aggregate word error rate for a given system, word error rates will be computed for pilots and controllers separately since these channels suffer from significantly different levels of distortion.

3.3 CALL-SIGN IDENTIFICATION

On air traffic control networks, there are typically multiple speakers and multiple listeners. A given transmission always has only a single speaker, but it may be directed at one or more listeners. Additionally,

successive transmissions will likely be spoken by different speakers and are likely to be directed at different sets of listeners. Speakers often identify themselves through the use of a call-sign, i.e., a sequence of words used as a unique identity for the airplane, pilot or controller. Some examples of call-signs are “Air France one fifty one heavy” and “Brussels approach”. Additionally, speakers often use call-signs to identify the identity of the airplane or controller position to whom they are directing the transmission. A given transmission may contain zero, one or multiple call-signs.

For the Call-Sign Identification (CSI) task the goal for a given system is to detect each occurrence of a spoken call-sign in each transmission. The system outputs a list of call-signs spoken during each transmission. This list may be empty if no call-signs are spoken. The structure of call-signs and the handling of partially spoken call-signs are discussed in detail in later sections of these guidelines.

The CSI task will be evaluated using precision/recall metrics which are standard for named-entity recognition tasks.

$$Precision = \frac{\#\ call\ signs\ correctly\ detected}{\#\ call\ signs\ hypothesized}$$

$$Recall = \frac{\#\ call\ signs\ correctly\ detected}{\#\ true\ call\ signs}$$

These metrics are used within the information retrieval community and are related to miss and false alarm measures (commonly used in detection tasks). A common operating point, the F-measure, that balances between these two measures will also be used:

$$F_1 = \frac{2(Precision * Recall)}{Precision + Recall}$$

As the F_1 measure assesses a single operating point, additional analysis may be done to characterize system performance across a range of potential operating conditions that may be more representative of real-world operation.

3.4 ENTITY CLUSTERING TASKS

Each ATC transmission can be associated with a speaker and an intended target. The Entity Clustering (EC) task is to identify the speaker/listener of each transmission. The identity can be an arbitrary string, but the system must use the same arbitrary string for representing the speaker in all transmissions produced by the speaker, and it must use a different arbitrary string for each speaker/listener.

To measure performance, the system speaker/listener identities are compared with the reference identities and an optimal entity mapping of reference speaker/listener identities to system speaker/listener identities is performed in a way that minimizes the EC error rate. The EC error rate is the sum of the EC errors divided by the total number of transmissions. For details about this process, refer to the description of NIST’s `md-eval.pl` [1].

3.4.1 Speaker Entity Identification

Each ATC transmission is spoken by a person representing a particular "entity". Pilots represent particular aircraft platforms (often identified by a call-sign). Controllers represent a particular ATC controller function (e.g., "approach", "tower", etc.). Often, a single pilot speaks all transmissions emanating from his aircraft. Sometimes, more than one pilot will speak transmissions from the same aircraft. The Speaker Entity Identification (SEI) task is to identify the entity represented by the speaker of the transmission either by call sign or a unique cluster-id represents the speaker. For each transmission, the system outputs an indication of whether the transmission was spoken by a pilot or by a controller.

Two SEI error measures are computed:

- Total SEI error is computed as the sum of SEI errors divided by the total number of transmissions. An error occurs if there is a pilot/controller confusion (i.e., system indicates pilot, reference indicates controller, or vice versa) or a pilot/pilot confusion (i.e., both system and reference indicate pilot, but the call signs don't match).
- SEI pilot/controller confusion error is computed by summing the SEI errors after ignoring the call-sign designations in both the system output and the reference annotation.

3.4.2 Listener Entity Identification

Each ATC transmission is directed to a particular "entity". Controllers direct transmissions to particular aircraft platforms (often identified by a call-sign). Pilots direct transmissions to a particular ATC controller function (e.g., "approach", "tower", etc.) The Listener Entity Identification (LEI) task is to identify the entity to which a transmission is directed. The system outputs an indication of whether the transmission was directed to a pilot or to a controller. For pilot-directed transmissions, the system must also specify the call-sign of the aircraft.

Some transmissions may be directed at all aircraft on the network. In such special cases, the system should indicate that the transmission was directed to all pilots.

Two LEI error measures are computed:

- Total LEI error is computed as the sum of LEI errors divided by the total number of transmissions. An error occurs if there is a pilot/controller confusion (i.e., system indicates pilot, reference indicates controller, or vice versa) or a pilot/pilot confusion (i.e., both system and reference indicate pilot, but the call signs don't match).
- LEI pilot/controller confusion error is computed by summing the LEI errors after ignoring the call-sign designations in both the system output and the reference annotation.

3.5 NATIVE/NON-NATIVE DETECTION (N3D)

In NATO operations soldiers may speak their native languages when communicating with forces of their own country but are likely to speak in English when communicating with forces of other countries. Most automatic speech processing systems perform better when they are informed of the language of the speech input, and many automatic speech processing systems perform better when they are told whether the speech being processed is native or non-native. Thus, it is valuable to be able to determine automatically the language and nativeness of a speech utterance.

In the nnMATC corpus, all speech is spoken in English. However, much of the English is spoken by non-native speakers. For this data set we define two accent detection/identification tasks:

1) The Native/Non-Native Detection (N3D) Task – determine the likelihood that the speaker of a transmission is a native speaker of English. The system outputs a score of arbitrary scale according to the convention that higher scores indicate greater likelihood that a transmission was spoken by a native speaker than lower scores.
2) The Accent Recognition (AR) Task – For a given transmission, determine the accent of the speaker from a closed set of possible accents. For each test message a system will generate a series of scores for each possible accent.

System performance is computed through the use of detection error trade-off (DET) curves. Transmissions are sorted by N3D/AR score. Probability of miss and false alarm are computed and plotted for all possible decision thresholds normalizing for detection priors.

3.6 PROCESSING OF DATA

The nnMATC corpus has been partitioned into three sets: **train**, **development** and **eval**. Final system performance will be evaluated using the **eval** set. Systems may use the training data to build acoustic models for ASR, generate long-term speaker models for clustering or any other training activity (setting thresholds, finding cohorts, etc.).

The development set has been set aside for system tuning and experiment prior to processing evaluation data. No tuning may be done using the evaluation set. As stated previously, data may be processed in any order in multiple passes. Sites are required to use the segmentation included in this corpus for submission of results. Files are provided with the appropriate reference segmentation information.

Sites wishing to perform evaluations of automatic segmentation and speech activity detector are free to use the provided files to assess segmentation performance, but for other tasks, submissions must make use of reference segmentation.

3.7 EVALUATION OF HUMAN BASELINE

In order to help establish a human baseline of performance for the tasks listed above, sites may determine results from human subjects following a similar protocol to the protocol described above. The same format and tasks will be available to human subjects for baseline analysis.

All annotations should be done using DGA's Transcriber tool. Empty Transcriber files, with segment marks will be provided as part of the nnMATC corpus. Please refer to the included README file for details. There are two differences between the human baseline process and the protocol described above.

1) *Training* – Human annotators should be given a set of training examples before starting the annotation of test set data. A set of examples are provided with this corpus for practice. Refer to the included README file for details.
2) *Time Limit* – Subjects will be asked to process the test data within a four hour time slot. Please inform human subjects that they are free to take rest breaks as needed.
3) *Call Sign Entities* – Human subjects should label speaker and listener entities with call signs in a normalized form so that scoring of entity clusters can be done automatically using the `md-eval.pl` tool as supplied by NIST [1].

3.8 REFERENCES

[1] NIST Speech Recognition Scoring Toolkit, latest version available at http://www.nist.gov/speech/tools."

Chapter 4 – EXPERIMENTAL RESULTS

4.1 INTRODUCTION

The presence of multilingual and non-native speech complicates the task faced by those who wish to apply automatic speech processing technology to military applications. Most automatic speech processing algorithms (e.g., speech recognition, speaker recognition, language recognition) operate in two phases. During the *training phase*, models are created from labeled training speech utterances using statistical method. During the *recognition phase*, the models built during training are used to hypothesize the words (or speaker, or language, etc.) of a new test utterance. Mismatched situations in which the training speech and test speech are spoken in different languages, or in which the training speech is spoken by native speakers but the test speech is spoken by non-native speakers, typically cause a degradation in performance of automatic speech processing systems vs. the performance obtained when only single-language, native speech is processed. Degradation can also be caused by the increased rate of disfluencies and other similar errors made by non-native speakers.

As part of the NATO Speech in Realistic Battlefield Environments Project, one of the speech corpora collected, labelled and distributed allows researchers to measure the effectiveness of speech processing systems on multilingual and non-native speech in a military battlefield environment. The nnMATC corpus, collected by Belgium, contains primarily English speech spoken by both native and non-native speakers (see Section 2.3.4).

The rest of this chapter of the report describes experiments performed on the nnMATC corpora. These experiments showed the impact of multilingual and non-native speech on speech recognition, speaker recognition and language recognition performance. In some cases, the effect on recognition accuracy was modest. In other cases, it was moderate or severe.

4.2 THE IMPACT OF BATTLEFIELD SPEECH

Multilingual and non-native speech impacts the performance of speech processing systems in a variety of ways.

4.2.1 Impact on Speech Recognition

Many hundreds of hours of transcribed training speech can be required to train acoustic and language models for speech recognition. When it is likely that non-native speech will be encountered during recognition, and when little non-native speech is available for training, a dilemma is faced: is it better to train on a small amount of non-native speech to avoid a training/testing mismatch but incur the penalty of poorly trained models, or is it better to use large amounts of native speech yielding well-trained, but mismatched, models. Depending on the circumstances, it may also be possible to adapt the well-trained native models to the non-native speech, to use acoustic models from one language to perform speech recognition in another, or to use multilingual acoustic models.

4.2.2 Impact on Speaker Recognition

Most conventional speaker recognition systems hypothesize the speaker of an utterance through extraction of features from the speech signal that is related to the speaker's vocal tract shape. To the extent that many

languages share a common set of sounds and to the extent that speakers of one language have vocal tracts that are generally similar to speakers of another language, one might predict *a priori* that the complexities of multilingual and non-native speech would have a less severe impact on speaker recognition vs. speech recognition. But other factors such as speaking rate, phone frequency of occurrence, hesitations, etc. could cause multilingual and non-native speaker recognition performance to degrade relative to performance on single-language, native speech.

4.2.3 Impact on Entity Recognition

Language identification systems use both acoustic and phonetic measurements to hypothesize the language of a speech utterance. Just as in the case of speech recognition, one would expect non-native speech to degrade performance vs. native speech because of acoustic and language-model mismatches.

4.3 SPEECH RECOGNITION EXPERIMENTS ON THE NNMATC CORPUS

The ATC speech corpus consists of audio data collected in Belgium during NATO pilot training exercises to and from various sites. The speech data includes interactions between pilot and tower during “taxi”, “transition” and “approach” flight sequences. The speech is heavily accented English in a high noise environment which poses a challenging problem for existing speech technologies.

A total of 12 hours of ATC data was collected by Stephane Pigeon of the Belgian Royal Military Academy. The data was then transcribed and segmented by hand so that it could be used to train and evaluate an automatic speech recognition system. Unfortunately, the transcription process was not complete and there were a few issues that needed to be addressed in order for the transcriptions to be useful for ASR.

One problem was an inconsistency in the lexicon of words used for transcribing the data. The heavily accented speech, foreign location names and ATC specific jargon made this a difficult task for the transcribers. The problem had to be addressed in order to create a pronunciation dictionary (a prerequisite for building an ASR system). Table 1 below gives some examples of the types of lexical errors that were found during this process.

Table 1: Lexical Transcription Errors

BROGEL	**BALEN**	**KOKSIJDE**
BROEGEL	BAHLEN	COOKSIDE
BROGO	BALEM	COXSIDE
BROHO	BALLEM	
BROKEL		
BRUGHEL		

The other more significant problem (and the one which required the most work to address) was the manual segmentation. The time alignments provided with the manual transcriptions were in many cases far from

accurate. The situation was addressed by training a small “boot-strap” recognizer on a trusted sub-set of the data, and then using it to decode the remaining data set. Then the transcriptions generated by the recognizer were compared to the manual transcriptions and a segmentation was created using the recognizer time alignments when the two transcriptions matched at an utterance boundary. The rough segmentation created by this process was used to train a new model and the process was repeated. The overall result was to bring the error rate down from 63.3% to 52.4% (on a small subset) in addition to creating a useful segmentation of the remaining data.

The 12 hours of ATC data was partitioned into approximately 10 hours of training data with the remaining data split into two test sets (see Table 2).

Table 2: ATC Data Partitions

Subset	Duration
Train	9.8 hours
Dev	40 minutes
Test	26 minutes

4.3.1 Evaluation using HTK and Sphinx Recognizers

Speech recognition has developed rapidly in the last few years. However, robustness is one of the main attributes that ASR systems lack. Current systems are still far from performing well under noisy environments (such as inside a car, out in the street, etc.). Here we present the results obtained when using Sphinx and HTK to decode the nnMATC corpus. Both of them are state-of-the-art HMM-based systems and are freely available.

4.3.1.1 Sphinx

The latest version of Sphinx 3 available (namely 3.6.3) has been used to perform our experiments [1]. It is composed of two modules: SphinxTrain [2] and Sphinx decoder [3]. The former can be used to create and train HMM based acoustic models. It also includes a front-end analysis module, which can be used to calculate the MFCCs from any given signal. The latter carries out the decoding of any given speech utterance, for which it uses the acoustic models trained by SphinxTrain.

4.3.1.2 SphinxTrain

Every ASR system needs to learn about the sounds that need to be decoded. First of all, the speech signal needs to be represented in a way that its phonetic properties are emphasized. In Sphinx, this is done by means of wave2feat, which computes the MFCCs from any given speech signal. Dynamic features (such as delta and delta-delta) are also automatically used, but they are not user configurable.

After the feature files are obtained, they can be used to train the acoustic models. Sphinx uses HMMs to characterize all the triphones that appear in the training corpus. Of course, things will become unmanageable if we assign a different HMM to every single triphone. In order to solve this problem, state sharing is applied. Similar states from different HMMs are grouped into a set, which is called *senone*. All the states belonging to a

senone share the same probability distribution. The amount of tying applied is user defined. The sphinx_train.cfg file sets the value of many of the training parameters, such as the number of states per HMM, the number of Gaussian mixtures used to model each senone, etc.

4.3.1.3 Force Alignment

In languages like English it is very common to find that the same word can be pronounced in several different ways. The dictionary file in Sphinx is allowed to have several entries for the same word. However, for the system to work properly, the transcription file must state which pronunciation alternative is used for each word. Sphinx provides a way to do this automatically, which is called forced alignment. We have used it to modify the transcription files provided by the MIT.

In order to use the forced alignment module, context independent models are trained using the original transcription (i.e., triphones are not created and only monophones are created at this step). Once these models are trained, the forced alignment script can be used to create a new set of transcription files. These will show, after each word, the pronunciation alternative used. Finally, these corrected transcriptions must be used to re-train the system. Of course, this time triphones will be created and they will be trained using the training corpus.

4.3.1.4 Sphinx 3 Decoder

The Sphinx decoder was used to recognize the test corpus. It uses models trained in prior steps and computes the most probable sequence of words associated with each utterance. To keep things manageable, the less probable HMMs are continuously being discarded during decoding. This approach is called pruning. The amount of pruning applied, among many others parameters, can be configured by means of the sphinx_decode.cfg file, as was done in SphinxTrain. After it ends, the user gets a hypothesis about what is said in each utterance of the test corpus. In order to get a measure of the performance of our system, this hypothesis can be compared to the transcription file. We have used the sclite for this task, which is part of the NIST's scoring package. It gets both the transcription (or reference) and hypothesis files as inputs, compares them, and calculates the *word error rate* achieved by the decoder.

4.3.1.5 Language Models

Speech recognition can be seen as the task of estimating a maximum a posteriori probability. Given any set of feature files, the system must obtain the most probable sequence of words associated to it. Any sequence of words may be generated by the acoustic models, even some that don't make sense, contain grammatical errors, etc.

The purpose of the *language model* is to make effective use of linguistic constraints when computing the probability of the different possible word sequences. For this task, we have used a trigram based language model, which has been created using the CMU-Cambridge Statistical Language Modeling Toolkit. These models assume that the probability of any word in the sequence depends only on the previous two words. In order to properly describe the kind of language used by both the pilots and the air controllers, our model has been created from the transcription for both the train and test utterances.

4.3.1.6 Results Obtained

The best WER value that we have been able to achieve is 36.2% when decoding the test corpus. To obtain this value, we trained our acoustic models keeping in mind that a word may be pronounced in several different ways. The transcription of every utterance in the training corpus was modified using the forced aligner, and then

triphone HMMs models were created from this data. Finally, these models were used to decode the test corpus. Both training and decoding configuration parameters and their values are described in [4]. It should be noted that if the original transcriptions are used (i.e., those not forced aligned), then the WER value dramatically increases to about 45%.

4.3.1.7 HTK

HTK [5] has also been used to decode the nnMATC database. The best WER value achieved is 59.60%. We tried to configure HTK using the very same parameters described for Sphinx, so as to be able to benchmark the performance of both systems. However, two problems arose. First of all, we were not able to create and train triphones in HTK, so models were created only for monophones. Secondly, HTK doesn't seem to be able to handle trigram language models, a bigram based language model was used instead.

4.3.2 Evaluation using DGA Recognizer

4.3.2.1 Experiment Descriptions

4.3.2.1.1 Pre-Processing Step

In order to fit the DGA ASR (Automatic Speech Recognition) system input requirement, the audio data sampling rate was converted to 8 kHz.

4.3.2.1.2 System 1

The first system tested was the DGA ASR English CTS (Conversational Telephone Speech) system without any adaptation or modification.

The different steps of that ASR system consist of:

- Feature extraction of the audio signal;
- Speech detection;
- Speaker segmentation and sex recognition;
- First decoding using a 3-gram language model and a 48 phone set;
- Vocal track length normalisation to reduce inter-speaker differences;
- Phone set pruning to 40 phones to ease speaker adaptation;
- Speaker adaptation of the acoustic model;
- Second decoding using a 3-gram language model and the 40 phone set; and
- Re-scoring using a 4-gram language model and the 40 phone set.

That baseline system obtained a 105.5 % WER (Word Error Rate) result on the dev set.

4.3.2.1.3 System 2

System 2 was a modification of system 1. Its language model was replaced by a 3-gram model built from the nnMATC train set thanks to the HTK toolkit and was provided by the MIT LL.

Also, its vocabulary was replaced by a vocabulary built (and provided) by MIT LL from the nnMATC train set.

Due to the fact that there is no BEEP model in the DGA system phone set, the [BEEP] annotations were removed from the reference annotation files in the test set. That removal reduced the WER by 0.5 % absolute on the dev set.

Also, it appeared that the filler words (ex: um, oh, uh, ah, ooh,…) used in the reference annotation files were not well normalized and consequently were causing errors. Consequently, they were also removed from the test set and that removal reduced the WER by 0.7 % absolute.

Thanks to those modifications, system 2 got a 67.5 % WER result on the dev set.

4.3.2.1.4 System 3

Basically, system 3 was the same as system 2 except that its language model was a 3-gram for the decoding phase and a 4-gram for the rescoring phase. Those models were built from the training data using the SRI Language Model Toolkit.

Thanks to that refined language model, system 3 got a 66.5% WER on the dev set which is a 1.2% absolute improvement compared to system 2.

4.3.2.1.5 System 4

Basically, system 4 was the same than system 3 except that the vocabulary used was taken from the DGA baseline system (system 1) which wasn't obtained from the nnMATC training data but from CTS data.

Furthermore, in order to reduce the OOVs (Out Of Vocabulary) words which are the words that are present in the nnMATC database but that are not known by the ASR system, a post-process replacement of the most frequent OOVs words in the database was applied on the ASR outputs. Those replacements reduced the WER by 1.0 % absolute.

Thanks to those modifications system 4 obtained a 65.2 % WER result on the dev set, which is a 1.3 % absolute improvement, compared to system 3.

4.3.2.1.6 Rover

The aim of that experiment was to fuse the 2 best ASR systems built (Sys 3 and Sys 4) which only differed in terms of vocabulary to check to what extent they were complementary.

The tool used to do that ROVER experiment was Sclite, and the alignment method used was the 'oracle' one, which means that Sclite always chose the best output of the fused ASR systems.

The score of that ROVER experiment is 58.0 % WER on the dev set which is 7.2 % better than the best single system.

4.3.2.2 Results Summary

Table 3 contains a summary of results obtained by the different systems when scored on the dev and test set.

Table 3: DGA Word Error Rates on Test Partitions

ASR Systems	Dev Set	Test Set
Sys 1: DGA baseline system	105.5 %	93.9 %
Sys 2: MIT lm + MIT voc	67.5 %	60.3 %
Sys 3: nnMATC lm + MIT voc	66.5 %	58.7 %
Sys 4: nnMATC lm + DGA voc	65.2 %	58.5 %
Rover (Sys 3 + Sys 4)	58.0 %	50.1 %

From this table, it appears that on the test set the systems are better than on the dev set. The reason for such differences is that in the test set annotation files, some segments contained speech that was not annotated as speech segments. Those segments resulted in a lot of insertion word errors when evaluated with Sclite. In order to prevent those segments penalizing the scores, DGA decided to remove them from the test set. As that work of excluding non-annotated speech segments was only done on the test set, the ASR systems scored better on that set.

4.3.2.3 Improvements

As the present experiments were only focused on language models and vocabulary aspects, an improvement would be to build a new acoustic model from the nnMATC train set. More specifically, as pilots and controllers speech are quite different acoustically and lexically, it might be interesting to build a segmentation model in order to differentiate them and then to build models (acoustic and language) for each of them.

Another improvement, which is very classical in speech processing, would simply consist in increasing the quantity of training data.

4.3.3 Evaluation using VOGON Recognizer

The ASR system used in these experiments is the Lincoln Labs “Vogon” system using non-cross-word triphone state clustered Gaussian mixtures (with a maximum of 12 Gaussians per state cluster). The system used gender independent models without speaker adaptation. A Vogon system was trained with 2048 state clusters and a closed vocabulary trigram model was created from the training data. The performance of the ASR system was between 52% and 54% WER on the two test sets (see Table 4).

Table 4: VOGON Word Error Rate on Test Partitions

Test Set	WER
Dev	52.7
Test	53.4

There are many ways in which the ASR system could be improved including the use of separate “pilot” and “tower” channel modeling, the incorporation of data from the Green Flag and FAA corpora, the use of speaker adaptation (not yet a feature of Vogon) or the use of an ATC specific grammar. The data also presents an opportunity for evaluating other technologies like speaker identification or “listener” identification, or closed set accent identification. While the initial word error rate is high for a direct ATC transcription task, the ASR transcriptions (or lattices) can be used for other token-based technologies.

4.4 CONCLUSIONS

A variety of experiments measuring the impact of multilingual and non-native speech on automatic speech processing accuracy have been performed. The results vary depending on the type of technology employed, the way in which the data are used, and the experimental methodology. Generally, however, we see that speech-processing performance degrades somewhat as we move from single-language, native applications to multi-lingual, non-native applications. Research efforts seeking to close this gap are underway at many sites worldwide.

4.5 REFERENCES

[1] Evandro Gouvea. Learning to use the CMU Sphinx Automatic Speech Recognition System. Carnegie Mellon University. http://cmusphinx.org/tutorial.html.

[2] Rita Singh. The SphinxTrain Manual. Carnegie Mellon University. http://www.speech.cs.cmu.edu/sphinxman/fr4.html.

[3] Mosur K. Ravishankar. Sphinx-3 s3.6 Decoder. Carnegie Mellon University. http://cmusphinx.sourceforge.net/sphinx3/s3_description.html.

[4] Francisco Sevilla, Raul Mohedano. Evaluation of the NATO IST031/RTG013 Non-Native Military Air Traffic Control (nnMATC) corpus using HTK and Sphinx recognizers (Internal Report). CIDA, Spain, 2006.

[5] Steve Young et al. The HTK Book (for HTK version 3.3). Cambridge University Engineering Department, 2005. http://htk.eng.cam.ac.uk.

Chapter 5 – RECOMMENDATIONS AND CONCLUSIONS

The field of military communications requires the integrated use of speech technology for command, control, and communications. In addition for multinational environments, it is necessary for a wide range of protocols from participating countries to be integrated together for safe and effective operations. Speech technology offers the promise of more direct and effective communications, verification of personnel, and allowing operators to have seamless access to information. Previous projects by this group [1] have shown that the problem of non-native speech raises a serious obstacle for the transition of commercial off-the-shelf (COTS) speech technology for speaker recognition, speaker verification, synthesis, and coding. Studies conducted as part of this project by participating NATO laboratories and reported here suggest that performance of COTS speech technology is degraded even more when that non-native speaker is in a stressful, noisy environment characteristic of most military environments. Advances in basic research to address this problem have not kept up with the demand for more widespread application of speech technology. It is hoped that this report will serve to focus the speech community on the important issue of speech and language variability due to non-native speech in battlefield environments. Databases collected during this study have been distributed to all participating NATO countries and some databases are available in CD-ROM format for those interested. Below we summarize the main finding and recommendations.

1) Military operations are often conducted in which multi-national coalition partners must communicate in a non-native language. These conditions are known to cause problems especially in stressful, noisy military environments.

2) These factors are detrimental to the effectiveness of communications in general, as well as to the performance of communications equipment and weapons systems equipped with vocal interfaces (e.g., advanced cockpits, command, control, and communications systems) trained for the native language.

3) Commercial off-the-shelf speech recognition systems are not yet able to address the wide variability associated with a non-native speaker.

4) Progress in the field of military-based speech technology has been restricted due to the lack of availability of databases of non-native speech in a military communications scenario.

5) It is certain that in the future it will be necessary to improve the coordination and effectiveness of multi-national military forces. The need therefore exists for planned simulations and exercises requiring coordinated emergency and/or emergency personnel using a wide range of speech technology. Such settings will have to address effective communications between multi-national forces using the same speech systems.

6) The success of the four-year effort by IST-031/RTG-013 has underlined the necessity to further invest coordinated international effort to support NATO interests in understanding speech production and perception and our ability to implement speech systems that are robust to the realities of everyday military speech.

5.1 REFERENCES

[1] “Implications of Multilingual Interoperability of Speech Technology for Military Use”, RTO-TR-IST-011, September 2004. ftp://ftp.rta.nato.int/PubFullText/RTO/TR/RTO-TR-IST-011/TR-IST-011-02.pdf.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's References	3. Further Reference	4. Security Classification of Document
	RTO-TR-IST-031 AC/323(IST-031)TP/57	ISBN 978-92-837-0060-9	UNCLASSIFIED/ UNLIMITED

5. Originator

Research and Technology Organisation
North Atlantic Treaty Organisation
BP 25, F-92201 Neuilly-sur-Seine Cedex, France

6. Title

Speech Processing in Realistic Battlefield Environments

7. Presented at/Sponsored by

This Technical Report has been prepared as a result of a project on "Speech Processing Using Realistic Battlefield Data" for the RTO Information Systems Technology Panel (IST) by Task Group 013.

8. Author(s)/Editor(s)	9. Date
Multiple	April 2009
10. Author's/Editor's Address	**11. Pages**
Multiple	48

12. Distribution Statement

There are no restrictions on the distribution of this document. Information about the availability of this and other RTO unclassified publications is given on the back cover.

13. Keywords/Descriptors

Control equipment	Pattern recognition	Speech recognition
Human factors engineering	Reviews	Speech technology
Intelligibility	Sensitivity	Standardization
Knowledge bases	Situational awareness	Voice communication
Languages	Software engineering	Voice control
Methodology	Speaker recognition	Voice recognition
Multilingualism	Speech analysis	Word recognition

14. Abstract

This report summarizes the results of research conducted by IST RTG-013 to better understand, detect, and mitigate the effects of native and non-native speech produced in military battlefield environments. Speech data was collected in three representative conditions to foster this and future research. Descriptions of the databases are included in the report. In addition, experimental plans on how to use these databases to measure the performance of speech processing systems are provided. Results using several state-of-the-art speech recognition systems from various IST RTG-013 member countries are presented. Finally recommendations and impact on current and future multi-national operations are presented.

NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

BP 25
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int

DIFFUSION DES PUBLICATIONS

RTO NON CLASSIFIEES

Les publications de l'AGARD et de la RTO peuvent parfois être obtenues auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la RTO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus soit à titre personnel, soit au nom de votre organisation, sur la liste d'envoi.

Les publications de la RTO et de l'AGARD sont également en vente auprès des agences de vente indiquées ci-dessous.

Les demandes de documents RTO ou AGARD doivent comporter la dénomination « RTO » ou « AGARD » selon le cas, suivi du numéro de série. Des informations analogues, telles que le titre est la date de publication sont souhaitables.

Si vous souhaitez recevoir une notification électronique de la disponibilité des rapports de la RTO au fur et à mesure de leur publication, vous pouvez consulter notre site Web (www.rto.nato.int) et vous abonner à ce service.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Gorch-Fock-Straße 7, D-53229 Bonn

BELGIQUE
Royal High Institute for Defence – KHID/IRSD/RHID
Management of Scientific & Technological Research for Defence, National RTO Coordinator
Royal Military Academy – Campus Renaissance
Renaissancelaan 30, 1000 Bruxelles

CANADA
DSIGRD2 – Bibliothécaire des ressources du savoir
R et D pour la défense Canada
Ministère de la Défense nationale
305, rue Rideau, 9e étage
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Acquisition and Logistics Organization (DALO)
Lautrupbjerg 1-5, 2750 Ballerup

ESPAGNE
SDG TECEN / DGAM
C/ Arturo Soria 289
Madrid 28033

ETATS-UNIS
NASA Center for AeroSpace Information (CASI)
7115 Standard Drive
Hanover, MD 21076-1320

FRANCE
O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72, 92322 Châtillon Cedex

GRECE (Correspondant)
Defence Industry & Research General Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HONGRIE
Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
P O Box 26
H-1525 Budapest

ITALIE
General Secretariat of Defence and National Armaments Directorate
5th Department – Technological Research
Via XX Settembre 123
00187 Roma

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

PAYS-BAS
Royal Netherlands Military Academy Library
P.O. Box 90.002
4800 PA Breda

POLOGNE
Centralny Ośrodek Naukowej Informacji Wojskowej
Al. Jerozolimskie 97
00-909 Warszawa

PORTUGAL
Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide
P-2720 Amadora

REPUBLIQUE TCHEQUE
LOM PRAHA s. p.
o. z. VTÚLaPVO
Mladoboleslavská 944
PO Box 18
197 21 Praha 9

ROUMANIE
Romanian National Distribution Centre
Armaments Department
9-11, Drumul Taberei Street
Sector 6
061353, Bucharest

ROYAUME-UNI
Dstl Knowledge and Information Services
Building 247
Porton Down
Salisbury SP4 0JQ

SLOVENIE
Ministry of Defence
Central Registry for EU and NATO
Vojkova 55
1000 Ljubljana

TURQUIE
Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanliklar
Ankara

AGENCES DE VENTE

NASA Center for AeroSpace Information (CASI)
7115 Standard Drive
Hanover, MD 21076-1320
ETATS-UNIS

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
ROYAUME-UNI

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council Acquisitions
Montreal Road, Building M-55
Ottawa K1A 0S2, CANADA

Les demandes de documents RTO ou AGARD doivent comporter la dénomination « RTO » ou « AGARD » selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications RTO et AGARD figurent dans les journaux suivants :

Scientific and Technical Aerospace Reports (STAR)
STAR peut être consulté en ligne au localisateur de ressources uniformes (URL) suivant: http://www.sti.nasa.gov/Pubs/star/Star.html
STAR est édité par CASI dans le cadre du programme NASA d'information scientifique et technique (STI)
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
ETATS-UNIS

Government Reports Announcements & Index (GRA&I)
publié par le National Technical Information Service
Springfield
Virginia 2216
ETATS-UNIS
(accessible également en mode interactif dans la base de données bibliographiques en ligne du NTIS, et sur CD-ROM)

NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

BP 25

F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int

DISTRIBUTION OF UNCLASSIFIED RTO PUBLICATIONS

AGARD & RTO publications are sometimes available from the National Distribution Centres listed below. If you wish to receive all RTO reports, or just those relating to one or more specific RTO Panels, they may be willing to include you (or your Organisation) in their distribution.

RTO and AGARD reports may also be purchased from the Sales Agencies listed below.

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number. Collateral information such as title and publication date is desirable.

If you wish to receive electronic notification of RTO reports as they are published, please visit our website (www.rto.nato.int) from where you can register for this service.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Royal High Institute for Defence – KHID/IRSD/RHID
Management of Scientific & Technological Research for Defence, National RTO Coordinator
Royal Military Academy – Campus Renaissance
Renaissancelaan 30
1000 Brussels

CANADA
DRDKIM2 – Knowledge Resources Librarian
Defence R&D Canada
Department of National Defence
305 Rideau Street, 9th Floor
Ottawa, Ontario K1A 0K2

CZECH REPUBLIC
LOM PRAHA s. p.
o. z. VTÚLaPVO
Mladoboleslavská 944
PO Box 18
197 21 Praha 9

DENMARK
Danish Acquisition and Logistics Organization (DALO)
Lautrupbjerg 1-5
2750 Ballerup

FRANCE
O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72, 92322 Châtillon Cedex

GERMANY
Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Gorch-Fock-Straße 7
D-53229 Bonn

GREECE (Point of Contact)
Defence Industry & Research General Directorate
Research Directorate, Fakinos Base Camp
S.T.G. 1020
Holargos, Athens

HUNGARY
Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
P O Box 26
H-1525 Budapest

ITALY
General Secretariat of Defence and National Armaments Directorate
5th Department – Technological Research
Via XX Settembre 123
00187 Roma

LUXEMBOURG
See Belgium

NETHERLANDS
Royal Netherlands Military Academy Library
P.O. Box 90.002
4800 PA Breda

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

POLAND
Centralny Ośrodek Naukowej Informacji Wojskowej
Al. Jerozolimskie 97
00-909 Warszawa

PORTUGAL
Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide
P-2720 Amadora

ROMANIA
Romanian National Distribution Centre
Armaments Department
9-11, Drumul Taberei Street
Sector 6
061353, Bucharest

SLOVENIA
Ministry of Defence
Central Registry for EU and NATO
Vojkova 55
1000 Ljubljana

SPAIN
SDG TECEN / DGAM
C/ Arturo Soria 289
Madrid 28033

TURKEY
Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanliklar – Ankara

UNITED KINGDOM
Dstl Knowledge and Information Services
Building 247
Porton Down
Salisbury SP4 0JQ

UNITED STATES
NASA Center for AeroSpace Information (CASI)
7115 Standard Drive
Hanover, MD 21076-1320

SALES AGENCIES

NASA Center for AeroSpace Information (CASI)
7115 Standard Drive
Hanover, MD 21076-1320
UNITED STATES

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
UNITED KINGDOM

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council Acquisitions
Montreal Road, Building M-55
Ottawa K1A 0S2, CANADA

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of RTO and AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
STAR is available on-line at the following uniform resource locator: http://www.sti.nasa.gov/Pubs/star/Star.html
STAR is published by CASI for the NASA Scientific and Technical Information (STI) Program
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
UNITED STATES

Government Reports Announcements & Index (GRA&I)
published by the National Technical Information Service
Springfield
Virginia 2216
UNITED STATES
(also available online in the NTIS Bibliographic Database or on CD-ROM)

ISBN 978-92-837-0060-9