DESIGNING
HUMAN INTERFACE
IN
SPEECH TECHNOLOGY
DESIGNING
HUMAN INTERFACE
IN
SPEECH TECHNOLOGY

FANG CHEN
Interaction Design,
Department of Computer Science
Chalmers University of Technology
Göteborg University
Sweden

Springer
Dedication

This book is dedicated to my two lovely boys

Henry and David
Contents

Preface..............................................................................................................xv
Abbreviation.....................................................................................................xix
Acknowledgements..........................................................................................xxiii

1.  INTRODUCTION .............................................................................................. 1
  1.1 NEW TIME WITH NEW REQUIREMENT .................................................. 1
  1.2 THE ASR APPLICATIONS ................................................................. 2
      The History
      Analysis of the Consumer Application
      Using ASR to drive Ship, One Application Example
  1.3 INTERFACE DESIGN .................................................................................. 9
      Neuroscience and Cognitive Psychology
      Speech Interaction Studies
      Multidisciplinary Aspects
      Task-Oriented or Domain-Oriented Systems
  1.4 USABILITY AND USER-CENTERED DESIGN ..................................... 13
      Usability Issues
      Understanding the User
      About Design Guidelines
1.5 RESEARCH PHILOSOPHY .......................................................... 17
   I Ching Philosophy
   Research and Design Methodology
   Research Validity

1.6 CONCLUSION ........................................................................... 25

2. BASIC NEUROPSYCHOLOGY...................................................... 27

2.1 INTRODUCTION ........................................................................ 27

2.2 GENERAL ASPECTS OF NEUROSCIENCE .............................. 28
   Basic Neuron Structure and Signal Transfer
   The Principles of Brain Functioning
   Methodology of Neuropsychological Studies

2.3 PERCEPTION ............................................................................ 33
   The Sensory System
   Sensory Modality
   Cortical Connection and Function
   Perception and Recognition Theories

2.4 LANGUAGE ............................................................................... 39
   Neurology of Language
   Speech Perception
   Word Recognition
   Language Comprehension

2.5 LEARNING AND MEMORY ....................................................... 45
   Learning Theory
   Memory
   The Mechanism of Memory
   Working Memory
   Forgetting

2.6 CONCLUSION .......................................................................... 50

3. ATTENTION, WORKLOAD AND STRESS................................. 53

3.1 INTRODUCTION ....................................................................... 53

3.2 ATTENTION .............................................................................. 54
   Neuroscience Perspective
   Focusing Attention
   Dividing Attention

3.3 MULTIPLE-TASK PERFORMANCE ............................................ 58
   The Resource Concept
   The Efficiency of Multiple Task Performance
   Individual Differences

3.4 STRESS AND WORKLOAD ...................................................... 61
   Stress
   Workload
3.5 THE RELATIONSHIP BETWEEN STRESS AND WORKLOAD .................. 63
   Stressors Classification
   Stress: Its Cognitive Effects
   Stress: Its Physiological Effects
   Stress: Its Effects on Speech
   Fatigue

3.6 THE MEASUREMENT OF STRESS AND WORKLOAD ..................... 68
   The Measurement of Stress
   Workload Assessment
   Performance Measurement
   Subjective Rating Measures

3.7 PSYCHOPHYSIOLOGICAL MEASURES ...................................... 72
   Psychological Function Test
   Physiological Function Test

3.8 ENVIRONMENTAL STRESS .................................................... 78
   Acceleration
   Vibration
   Noise
   Auditory Distraction

3.9 WORKLOAD AND THE PRODUCTION OF SPEECH ............................. 83

3.10 ANALYSIS OF SPEECH UNDER STRESS .................................... 85
   Emotion and Cognition
   Speech Measures Indicating Workload
   Acoustic Analysis of Speech
   Improving ASR Performance

3.11 RESEARCH PROBLEMS ....................................................... 91

3.12 CONCLUSION ..................................................................... 94

4. DESIGN ANALYSIS .................................................................. 95

4.1 INTRODUCTION .................................................................... 95

4.2 INFORMATION PROCESSING THEORY ..................................... 97

4.3 THE ECOLOGICAL PERSPECTIVE ........................................ 98
   Ecological View of the Interface design

4.4 DISTRIBUTED COGNITION .................................................. 103

4.5 COGNITIVE SYSTEM ENGINEERING .................................... 104

4.6 WORK AND TASK ANALYSIS ............................................. 106
   Task Analysis
   Cognitive Task Analysis
   GOMS—A Cognitive Model
   Cognitive Work Analysis

4.7 INTERACTION DESIGN PROCESS ......................................... 115
   Human-System Interface
   Design Process
Designing human interface in speech technology

Interaction Design
4.8 Scenario-Based Design ................................................. 119
4.9 Discussion .................................................................. 120
4.10 Conclusion ................................................................. 122

5. Usability Design and Evaluation ................................. 123

5.1 Introduction ................................................................. 123
5.2 Different Design Approaches ...................................... 123
5.3 The Concept of Usability ................................................ 125
   Definitions
   Usability Design
   Usability Evaluation

5.4 Human Needs and Satisfaction .................................. 135
   Trust
   Pleasure

5.5 User-Centered Design .................................................. 141
   What is UCD Process?
   Planning the UCD Process
   Specifying the Context of Use
   User Analysis
   User Partnership
   Usability Requirements Analysis
   UCD Process in Practices

5.6 Social Technical Issue .................................................. 152
5.7 Adaptive and Intuitive User Interface ......................... 153
5.8 Usage-Centered Design ................................................ 155
   Creative Design
   Design Criteria and Process
   Ecological Approach

5.9 Universal Access .......................................................... 160
5.10 Ethnography Method for Contextual Interface Design ... 163
5.11 Conclusion ................................................................. 164

6. Human Factors in Speech Interface Design .................. 167

6.1 Introduction ............................................................... 167
6.2 The Unique Characteristics of Human Speech ............... 169
6.3 Human Speech Recognition System ............................. 171
   Automatic Speech Recognition System
   Speech Feature Analysis and Pattern Matching
   Speech Synthesis
   National Language Processing
   Language Modeling

6.4 Human Factors in Speech Technology ........................ 179
Designing human interface in speech technology

6.5 **Speech Input** .................................................................180
   Human Factors and NLP
   Flexibility of Vocabulary
   Vocabulary Design
   Accent
   Emotion

6.6 **Assessment and Evaluation** ........................................186

6.7 **Feedback Design** .......................................................188
   Classification of Feedback
   Modality of Feedback
   Textual Feedback or Symbolic Feedback

6.8 **Synthesized Speech Output** ........................................192
   Cognitive Factors
   Intelligibility
   Comprehension
   Emotion
   Social Aspects
   Evaluation

6.9 **Error Correction** .......................................................199
   Speech Recognition Error
   User Errors
   User Error Correction

6.10 **Syntax** .................................................................205

6.11 **Back-up and Reversion** ............................................206

6.12 **Human Verbal Behavior in Speech Input Systems** ..........208
   Expertise and Experience of the User
   The Evolutionary Aspects of Human Speech

6.13 **Multimodal Interaction System** .................................212
   Definitions
   Advantages of Multimodal Interface
   Design Questions
   Selection and Combination of Modalities
   Modality Interaction
   Modality for Error Correction
   Evaluation

6.14 **Conclusion** ............................................................224

7. **The Usability of Spoken Dialogue System Design** 225

7.1 **Introduction** ............................................................225

7.2 **The Attractive Business** .............................................226

7.3 **Ergonomic and Socio-Technical Issues** .......................228
   The User Analysis
   The Variance of Human Speech
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.4</td>
<td>SPEECH RECOGNITION ERROR</td>
<td>232</td>
</tr>
<tr>
<td>7.5</td>
<td>COGNITIVE AND EMOTIONAL ISSUE</td>
<td>234</td>
</tr>
<tr>
<td>7.6</td>
<td>AFFECTIVE COMMUNICATION</td>
<td>237</td>
</tr>
<tr>
<td>7.7</td>
<td>LIMITATIONS OF SUI</td>
<td>238</td>
</tr>
<tr>
<td>7.8</td>
<td>USABILITY EVALUATION</td>
<td>241</td>
</tr>
<tr>
<td>7.9</td>
<td>CONCLUSION</td>
<td>249</td>
</tr>
<tr>
<td>8.1</td>
<td>INTRODUCTION</td>
<td>251</td>
</tr>
<tr>
<td>8.2</td>
<td>IN-VEHICLE SPEECH INTERACTION SYSTEMS</td>
<td>256</td>
</tr>
<tr>
<td>8.3</td>
<td>THE COGNITIVE ASPECTS</td>
<td>266</td>
</tr>
<tr>
<td>8.4</td>
<td>USE OF CELLULAR PHONES AND DRIVING</td>
<td>272</td>
</tr>
<tr>
<td>8.5</td>
<td>USE OF IN-VEHICLE NAVIGATION</td>
<td>276</td>
</tr>
</tbody>
</table>
8.6 SYSTEM DESIGN AND EVALUATION .............................................. 281
   System Design
   System Evaluation/Assessment

8.7 FUTURE WORKS ........................................................................... 286
8.8 CONCLUSION .................................................................................. 287

9. SPEECH TECHNOLOGY IN MILITARY APPLICATION ...... 289

9.1 INTRODUCTION ............................................................................. 289
9.2 THE CATEGORIES IN MILITARY APPLICATIONS ...................... 291
   Command and Control
   Computers and Information Access
   Training
   Joint Force at Multinational Level

9.3 APPLICATION ANALYSIS .............................................................. 294
9.4 COMPARISON BETWEEN SPEECH INPUT AND MANUAL INPUT .... 297
   The Argumentation between Poock and Damper
   Effects of concurrent tasks on Direct Voice Input
   Voice Input and Concurrent Tracking Tasks

9.5 AVIATION APPLICATION .............................................................. 301
   The Effects from Stress
   Compare Pictorial and Speech Display
   Eye/Voice Mission Planning Interface (EVMPI) Model
   Application in Cockpit Fast Jet
   Battle Management System
   UAV Control Stations

9.6 ARMY APPLICATION ................................................................... 312
   Command and Control on Move (C2OTM)
   ASR Application in AFVs
   The Soldier's Computer
   Applications in Helicopter

9.7 AIR TRAFFIC CONTROL APPLICATION ................................... 316
   Training of Air Traffic Controllers
   Real Time Speech Gisting for ATC Application

9.8 NAVY APPLICATION .................................................................... 320
   Aircraft Carrier Flight Deck Control

9.9 SPACE APPLICATION ................................................................... 321

9.10 OTHER APPLICATIONS ............................................................... 322
   Computer Aid Training
   Aviation Weather Information
   Interface Design for Military Datasets

9.11 INTEGRATING SPEECH TECHNOLOGY INTO MILITARY SYSTEMS... 324
   The Selection of Suitable Function for Speech technology
Designing human interface in speech technology

Recognition Error, Coverage and Speed
Interface Design
Alternative and Parallel Control Interface
Innovative Spoken Dialogue Interface

9.12 CONCLUSION........................................................................................................... 330

References.................................................................................................................... 331
Index............................................................................................................................... 377
Preface

There is no question of the value of applying automatic speech recognition technology as one of the interaction tools between humans and different computational systems. There are many books on design standards and guidelines for different practical issues, such as Gibbon’s book *Handbook of Standards and Resources for Spoken Language System* (1997) and *Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology and Product Evaluation* (2000), Jurafsky’s *Speech and Language Processing* (2000), Bernsen, et al.’s book *Designing Interactive Speech System* (1998), and Balentine and Morgan’s book *How to Build a Speech Recognition Application – A style Guide for Telephony Dialogues* (2001), etc. Most of these books focus on the design of the voice itself. They provide certain solutions for specific dialogue design problems. Because humans are notoriously varied, human perception and human performance are complicated and difficult to predict.

It is not possible to separate speech behavior from their cognitive behavior in their daily. What humans want, their needs and the requirements from the interacting systems are changing all the time. Unfortunately, most of the research related to speech interaction system design focus on the voice itself and forgets about the human brain, the rest of the body, the human needs and the environment that will affect the way of their thinking, feeling and behavior. Differing from other books, this book focuses on understanding the user, the user’s requirements and behavior limitation. It focuses on the user’s perspective in the application context and environment. It considers on social-technical issues and working environment issues.
In this book, I will a give brief introduction of fundamental neuron science, cognitive theories on human mental behavior and cognitive engineering studies and try to integrate the knowledge from different cognitive and human factors disciplines into the speech interaction system design. I will discuss research methodologies based on the research results from human behavior in complicated systems, human-computer interaction and interface design. The usability concept and user-centered design process with its advantages and disadvantages will be discussed systematically.

During the preparation of this book, I strongly realized that there is very little research work in the human factor aspects of speech technology application studies. This book will emphasize the application, providing human behavior and cognitive knowledge and theories for a designer to analyze his tasks before, during and after the design and examples of human factors research methods, tests and evaluation enterprises across a range of systems and environments. Its orientation is toward breadth of coverage rather than in-depth treatment of a few issues or techniques. This book is not intended to provide any design guidelines or recommendations.

This book shall be very helpful for those who are interested in doing research work on speech-related human computer interaction design. It can also be very helpful for those who are going to design a speech interaction system, but get lost from many different guidelines and recommendations or do not know how to select and adjust those guidelines and recommendations into specific design, as this book provides the basic knowledge for the designers to reach any goals in design. For those who would like to make some innovative design with speech technology, but do not know how to start, this book may be helpful. This book can be useful for those who are interested in Human-Computer Interaction (HCI) in general and would like to have advanced knowledge in the area. This book can also be a good reference or course book for postgraduate students, researchers or engineers who have special interests in usability issues of speech-related interaction system design.

The present book will be divided into two parts. In part one, the basic knowledge and theories about human neuroscience and cognitive science, human behavior and different research ideas, theories and methodologies will be presented. It will provide design methodologies from a user-centered point of view to guide the design process to increase the usability of the products. It will cover the tests and evaluation methodologies from a usability point of view in variances of design stages. The human-factors related speech interface design studies in the literature are reviewed systematically in this part.

In part two, the book will focus on the application issue and interface design of speech technology in various domains such as telecommunication,
in-cars information systems and military applications. The purpose of this part is to give the guides for the application of cognitive theories and research methodologies to the speech-related interface design.
Abbreviations

ACC  Adaptive Cruise Control
ACT  Alternative Control Technology
AFVs  Armored Fight Vehicles
AH  abstraction hierarchy
AMS  Avionics Management System
ANN  artificial neural network
ASR  automatic speech recognition
ASW  anti-submarine warfare
ATC  air traffic control
ATT  Advanced and Transport Telemetric
AUIS  Adaptive User Interfaces
BP  blood pressure
C2OTM  Command and Control on Move
C3I  command, control, communications and intelligent
CAI  Computer-Assisted Instruction
CAVS  Center for Advanced Vehicular Systems
CFF  critical flicker frequency
CFG  context-free grammar
CLID  Cluster-Identification-Test
CNI  communication, navigation and identification
CNS  central nervous system
CSE  Cognitive System Engineering
CTA  cognitive task analysis
CUA  context of use analysis
CWA  cognitive work analysis
CWS  collision-warning system
DM  dialogue manager
DRT  diagnostic rhyme test
DVI  direct voice input
EEG  electroencephalogram
EID  ecological interface design
Designing human interface in speech technology

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>EOG</td>
<td>electrooculogram</td>
</tr>
<tr>
<td>ERP</td>
<td>event-related potential</td>
</tr>
<tr>
<td>EVMPI</td>
<td>Eye/Voice Mission Planning Interface</td>
</tr>
<tr>
<td>FAA</td>
<td>Federal Aviation Administration</td>
</tr>
<tr>
<td>fMRI</td>
<td>function MRI</td>
</tr>
<tr>
<td>GA</td>
<td>global array</td>
</tr>
<tr>
<td>GCS</td>
<td>ground control system</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Position System</td>
</tr>
<tr>
<td>GIDS</td>
<td>generic intelligent driver support</td>
</tr>
<tr>
<td>GOMS</td>
<td>goals, operators, methods selection roles</td>
</tr>
<tr>
<td>GUI</td>
<td>graphical user interface</td>
</tr>
<tr>
<td>HCI</td>
<td>human-computer interaction</td>
</tr>
<tr>
<td>HMM</td>
<td>Hidden Markov Model</td>
</tr>
<tr>
<td>HR</td>
<td>heart rate</td>
</tr>
<tr>
<td>HRV</td>
<td>heart rate variability</td>
</tr>
<tr>
<td>HTA</td>
<td>Hierarchical Task Analysis</td>
</tr>
<tr>
<td>HUD</td>
<td>head-up display</td>
</tr>
<tr>
<td>IBI</td>
<td>inter-beat interval</td>
</tr>
<tr>
<td>ICE</td>
<td>in-car entertainment</td>
</tr>
<tr>
<td>IP</td>
<td>information processing</td>
</tr>
<tr>
<td>ITS</td>
<td>Intelligent Transport Systems</td>
</tr>
<tr>
<td>IVCS</td>
<td>in-vehicle computer system</td>
</tr>
<tr>
<td>IVHS</td>
<td>Intelligent Vehicle Highway Systems</td>
</tr>
<tr>
<td>IVR</td>
<td>interactive voice response</td>
</tr>
<tr>
<td>KBB</td>
<td>knowledge-based behavior</td>
</tr>
<tr>
<td>LCD</td>
<td>liquid crystal display</td>
</tr>
<tr>
<td>LVCSR</td>
<td>large-vocabulary speaker-independent continuous</td>
</tr>
<tr>
<td>MCDS</td>
<td>Multifunction Cathode ray tube Display System</td>
</tr>
<tr>
<td>MEG</td>
<td>magnetonencephalogram</td>
</tr>
<tr>
<td>MHP</td>
<td>Model Human Processor</td>
</tr>
<tr>
<td>MRI</td>
<td>magnetic resonance imaging</td>
</tr>
<tr>
<td>MRT</td>
<td>modified rhyme test</td>
</tr>
<tr>
<td>NASA-TLX</td>
<td>NASA Task Load Index</td>
</tr>
<tr>
<td>NL</td>
<td>Natural Language</td>
</tr>
<tr>
<td>NLU</td>
<td>Natural Language Understanding</td>
</tr>
<tr>
<td>PET</td>
<td>positron emission tomography</td>
</tr>
<tr>
<td>PNS</td>
<td>peripheral nervous system</td>
</tr>
<tr>
<td>POG</td>
<td>point-of-gaze</td>
</tr>
<tr>
<td>PRF</td>
<td>performance-resources function</td>
</tr>
<tr>
<td>PRP</td>
<td>psychological refractory period</td>
</tr>
<tr>
<td>RBB</td>
<td>rule-based behavior</td>
</tr>
<tr>
<td>RSA</td>
<td>respiratory sinus arrhythmia</td>
</tr>
<tr>
<td>RTI</td>
<td>Road Transport Informatics</td>
</tr>
<tr>
<td>SAACL</td>
<td>Stress Arousal Checklist</td>
</tr>
<tr>
<td>SBB</td>
<td>skill-based behavior</td>
</tr>
<tr>
<td>S-C-R</td>
<td>stimulus-central processor-response</td>
</tr>
<tr>
<td>SIAM</td>
<td>speech interface assessment method</td>
</tr>
<tr>
<td>SQUID</td>
<td>superconducting quantum interference device</td>
</tr>
<tr>
<td>SRK</td>
<td>skill, rule, knowledge</td>
</tr>
<tr>
<td>SSM</td>
<td>soft-system methodology</td>
</tr>
</tbody>
</table>
Designing human interface in speech technology

SUI .................................................. speech user interface
SWAT ............................................. subjective workload assessment techniques
TEO ................................................ teager energy operator
TICS ............................................... Transport Information and Control Systems
TOT ............................................... Time on Task
TTS ............................................... text-to-speech
UAVs .............................................. Unmanned Aerial Vehicles
UCD ............................................... user-centered design
UWOT ............................................. User Words on Task
WER ............................................... Word Error Rate
Acknowledgements

I would like to thank Professor Rinzou Ebukuro, Certified Ergonomist, JES, Japan, helped me prepare the material and the texts for Chapter 1. Lay Lin Pow, Zhan Fu and Krishna Prasad from the School of Mechanical and Production Engineering Nanyang Technological University, Singapore, helped me prepare the material and the texts for Chapters 7 and 8.
Chapter 1

INTRODUCTION

1.1 NEW TIME WITH NEW REQUIREMENT

The fast-developing information technology with its quickly growing volume of globally available network-based information systems has made a strong impact on the changes of an individual’s life and working environment and the organization and social structure in many different aspects. The highly flexible, customer-oriented and network-based modern life requires the human-system interaction to be suitable for a broad range of users to carry out the knowledge-intensive and creative work under different contexts and various cultural backgrounds. Thus the requirement for the interface design is changing due to the differences of environment, organization, users, tasks and technology development. Nowadays, users are facing an overwhelming complexity and diversity of content, function, and interaction methods. Talking to the machine is long cherished human desire, and it is regarded as one of the most natural human-system interaction tools in the socio-technical systems context as repeatedly told in stories of science fiction.

In the later ‘90s, speech technology made its quick development. This is due to the extensive use of statistical learning algorithm, and the availability of a number of large collections of speech and text samples. The possibility of the high accuracy rate of automatic speech recognition (ASR) technology and the relative robust, speaker-independent, spontaneous (or continuous) spoken dialogue system, and humanized synthetic speech, together with the dramatically increase of computer speed, storage and bandwidth capacity, offered the new possibility for human beings to communicate with computer information systems through speech.
The development of speech technology creates the new challenge to the interface designer. As Bernsen (2000; 2001) pointed out, the speech field is no longer separate from many other fields of research, when it works for a broad range of application purpose, the human factor skills as well as system integration skills, among other skills are needed. According to his opinion, in the coming few years, the application of speech technology will still be task-orientated. So “the system must be carefully crafted to fit human behavior in the task domain in order to work at all” (2000). This requires deep understanding of humans, human behavior, the task and the system that are going to perform under the context of society, organization and performance environment.

The application of speech technology to different industrial fields has over a 30 years’ history. However, there are more cases of failure than success in the past years. Even up to now, users are still hesitating to use speech interaction systems; very few of them can reach the users’ demands of the system performance. People often blame the failure of the technology itself, because the automatic speech recognition (ASR) technology has insufficient accuracy. Researchers and engineers are still looking for better solutions to increase the recognition accuracy. Based on the accumulated application experienced from the past years, we should realize that the failure of the recognition accuracy probably is just part of the reason of the unsuccessful application of the technology. Some of the problems have originated mainly from lack of the understanding of the end users, their tasks, their working environments and their requirements. There is little systematic investigation about the reasons of failures on the applications, other than the misrecognition that has been reported in this field. Actually, the ASR system, based on the statistical calculation of voice spectrum matching, can hardly reach the human-like speech recognition performance, as there are so many factors that can affect human speech, therefore making varied voice from any possible voice samples. To be able to have the successful design of a speech-related interaction system, one needs to understand the constraints of the technology, to find appropriate and adequate applications and to make the proper interaction design that fits the user’s nature in cognition and performance. These are typically the weak points, and very few experts could explain it to the prospects.

1.2 THE ASR APPLICATIONS

In this section, we take a brief view of automatic speech recognition technology applications and outline the progress in the US and Japan in history. It is found that there are other factors besides the technology itself
that contribute to either the success or failure of the application. These factors can be: the problems of finding appropriate speech applications; lack of communications among cognitive researchers, application engineers, management and the end users; and further solid human factor research and development for applications.

1.2.1 The History

Dr. T., B. Martin, RCA Corporation originated the application history of speech input in the United States in 1969 with the installation of a speech recognition system at the Pittsburgh Bulk Mail Center of USPS for destination input to the mailbag-sorting system. In May 1970, Dr. Martin established the well-known Threshold Technology Inc. (TTI) for speech business development. TTI was listed in the stock market NASDAQ in 1972. Then he established EMI Threshold Limited in the United Kingdom in 1973 for development of its European market.

A TTI-installed voice input for sorting destinations and inspection data at United Air Lines, a luggage sorting system at Trans World Air Line in 1973, and after that at least an 11-systems operation were confirmed in the market by a well-known speech researcher (NEC's internal report, internally issued by Mr. Y. KATO 1974, not published). Since then the application field has been enlarged to different areas such as the inspection data input of Pull-Ring Can, automatic assembly line, inspections at delivery reception site, physical distribution centers, automobile final-inspection data input for quality controls at General Motors, cash registers at convenience stores, warehouse storage controls, cockpit simulations, air traffic control, supporting equipment for physically handicapped persons, map drawings, depth soundings and so forth (Martin, 1976). Most of them were for industrial applications. There were over 950 installations in the U.S. market around 1977, according to the sales experts of TTI.

NEC Corporation started up the speech business in Japan in 1972. It developed the well-known DP matching method (Sakoe, 1971). The system allows continuously spoken word input instead of discrete input that was commonplace for industrial field applications at that time. The first experimental DP model for field applications was completed in 1976. A trend of field introduction of the speech technology began with the sorting input in 1977 and 1978 (Abe, 1979; Suzuki, 1980) and then spread out to the other fields, such as meat inspection data input, steel sheet inspections, steel coil field storage controls, crane remote control input, cathode ray tube inspections, automobile final inspections and so forth, just as the application development in the US. There is an interesting survey result on DP-100 installations and inquiry from prospects for introduction in 1982 as shown in
Figure 1-1 (Ebukuro, 1984a; 1984b). Many other manufacturers entered into this market one after another in that age but many of them disappeared soon after, due to the fact that the business scale was too small and the required investment for the research and development was too much. There was very little systematic investigation about why manufacturers gave up the already installed voice input system. This probably was regarded as not only the factory’s secret, but also the lack of research and developments on the application basics.

Speech recognition for PC input was introduced to the Japanese market by the Japanese-affiliated company of a well-known U.S. manufacturer in around 1997 and was immediately followed by some Japanese manufacturers. Many of the main suppliers have withdrawn their products from this consumer market in Japan. The interesting phenomenon is that the sales talks in trite expressions this time were almost the same as the early ‘80s. Examples of sales talks from for vendors are summarized in Table 1-1.

Table 1-1. Group of examples of sales talks in trite expressions

- Japanese word recognition has improved input accuracy.
- You can use it by only 1-min. adjustments.
- Simple/Impressive.
- Voice input makes PC input easier.
- New century by voice.
- You can do chat with PC in a dialog sense.
1.2.2 Analysis of the Consumer Applications

The report of the analysis of American market status was published by the ZD Net News December 22, 2001. The main points in this report can be summarized as:

a) The real reason of why industrial and market statuses did not fit with the expectation of the application possibilities was not clear.
b) The real market scale was not clear.
c) Speech experts can rarely explain real state-of-speech application technology; e.g., product adaptability and adjustability to the applications and suitably applicable areas and site to the prospects.
d) Hand-free speech application developments for consumer applications may still require some more time.
e) Popularization of speech input technology may take time.
f) PC users (consumers) may abandon the speech products.
g) Only call-centers and supply chains look to gain a profit.

The situation of the Japanese market was not the exception as well. Why had those brilliant installations like the sorting centers eventually disappeared? What would happen to the other potential applications such as to consumer products and system solutions application? Although there are numerous reasons for disappearing and thriving, one of the main reasons from the view of the application basics is that we ignore the human factors study related to human speech in these working settings. The managements were interested in decreasing the workloads and improving the working environments, while the designers had too little knowledge about the human speech performance under the task behavior nature since many of them have disregarded the importance of human factors on the human behavior in the system operations.

The speech input is simple and impressive; voice input makes input tasks easier since the speech is human nature. These are true but not always and not everywhere. A comparison between the disappearing and thriving of ASR Systems in industrial applications is given in Table 1-2. It is summarized from the ergonomic point of view and shows that the advantages and disadvantages of the application depend on the characteristics of the tasks and the users. This historical fact showed us again the truth that an advanced technology can be hazardous if we do not apply it properly.

1 http://www.zdnet.co.jp/news/0010/20voice
Introduction

Researchers and development engineers may tell lofty ideas on speech input prospective aspects with their deep philosophical and practical experiences standing firm on the rolling deck. These may be based on a lot of premises that cannot tell at once and are sometimes implicitly hidden in explanations. We, vendors and the users, including the managements and the end users, must know and understand this nature of engineers. The application study may fill up this type of communication gap. The lack of communications and understanding among designers, vendors and users could make the situation worse.

Table 1-2. Differences between disappearing and thriving

<table>
<thead>
<tr>
<th>Difference</th>
<th>Disappearing (Luggage Sorting Input)</th>
<th>Thriving (Meat Inspection)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operators Vocabulary</td>
<td>Unskilled workers</td>
<td>Qualified personnel</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>Some tens of words for sorting</td>
<td>Not so many words for</td>
</tr>
<tr>
<td></td>
<td>destinations</td>
<td>inspection results</td>
</tr>
<tr>
<td>Inputs</td>
<td>Destinations that they are obliged</td>
<td>Familiar input words that</td>
</tr>
<tr>
<td></td>
<td>to read</td>
<td>are spoken under judgment</td>
</tr>
<tr>
<td>Work load and</td>
<td>Requirement of higher input speed</td>
<td>The speed follows on the</td>
</tr>
<tr>
<td>working environment</td>
<td>Facing the objects in finding</td>
<td>auctioning speed</td>
</tr>
<tr>
<td></td>
<td>destinations that requires effort to</td>
<td>Automatically fed hanged</td>
</tr>
<tr>
<td></td>
<td>find and read out the destinations</td>
<td>meat that requires little</td>
</tr>
<tr>
<td></td>
<td>Fixed position</td>
<td>effort to facing and finding.</td>
</tr>
<tr>
<td></td>
<td>Noises from higher-speed running</td>
<td>Can walk around</td>
</tr>
<tr>
<td></td>
<td>machine</td>
<td>Noises from slower running</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Machine</td>
</tr>
</tbody>
</table>

How should we improve the communications among different partners in the business chain such as researchers, technology engineers, speech analysts, vendors, industrial managements and the end users? The principle that a better understanding of the user, the task characteristics, user requirements, the working environment and the organization may bring all the partners in the business chain together to communicate and understand each other.

A well-known successful case of this kind of sophisticated technology in Japan is the OCR (Optical Character Reader) applications for postal code readings. The engineers and business developers have worked closely together with the postal administration, which followed the introduction of the principle of management on the errors and took the socio-technical systems approach. They have had complete understanding of the different aspects related to the application of the ASR technology and have developed a smart method of system introductions to the user. The printing bureau of
finance administration Japan has had applied this principle and had the successful installations of a sophisticated printing quality assurances system.

1.2.3 Using ASR to Drive Ship, One Application Example

The Ship Building Division developed the speech input sailing system called Advanced Navigation Support System (ANSS) with Voice Control and Guidance by Mitsubishi Heavy Industries Ltd. in September 1997 in applying NEC Semi-syllable speech input system DS-2000 and a voice response system (Fukuto, 1998). The system allows a diagonal marine ship operation applying a speech response system under the assistance of the Navigation Support System. Since then, nearly ten marine ships have been equipped with this system for their one-man bridge operations.

The first system was successfully installed on the coastal service LPG carrier called the SIN-PUROPA (propane) MARU (Gross Tonnage, 749t, Deadweight, 849t) owned by Kyowa Marin Transportation. The first technical meeting on the successful completion of the trial operation of the SIN-PROPAN MARU was held in Tokyo on July 1, 1998. It was reported that the error rate of voice input was first 8% to 9% or more, and finally it was reduced to 2% (by applying noise-canceling method and fine tunings). The training period for three navigators who had not been preliminary acquainted with the ship operation was only three days at the primary stage of ship operation, including speech input trainings. Two captains, who were firstly very skeptical about the speech navigations, made the following comments about speech navigations:

a) The voice operation was very easy even in the congested waters and narrower route. It was enjoyable, just like playing games.

b) We can have extra time to do the things other than the formal operations, e.g., give an earlier operational direction to the controller than the past navigations.

c) We have had the feeling of solidarity with the ship controller.

d) The navigator can concentrate his effort to watching the route for sailings, acquisition of information for navigation, radar range switching, and so forth, as the eyes and hands were free from machine operation.

It is very interesting to look at the implementation process of this successful achievement. It is again telling us how important it is to maintain the ergonomic and socio-technical systems concept in mind, such as the deep understanding of the operators, behavioral psychology, on their tasks requirement and good communication among different partners in the chain.
The statistics of collision on Marine ship sailing (Report on Marine Accident 2002, by Japan marine accident Inquiry Agency) showed that collision from insufficient watching was 56.8% in 2002 in Japan. The remainders are illegal sailing (18.8%), ignorance of signal (9.5%) and the others (14.9%). It shows the significant influence of safety sailing by watching at the bridge. Continuously observing traffic situations and various displays working with the navigation systems at the bridge are heavy mental and physical burdens for a captain. Speech technology introduced in the communication by the ANSS in operations has effectively reduced the burdens by keeping eyes and hands free. The further feelings of intimations to the machine system are confirmed. The expected effectiveness for the protection of collisions (Fukuto, 1998) is also confirmed.

Safety issues caused by the loss of situation awareness and the potential collision due to the performance constraints and the mental workload of operators can be very sensitive. Unlike the parcel sorting errors, a trivial error in operation may sometimes lead to significant marine disasters if the captain cannot take countermeasures to the critical or dangerous situations when an error has occurred.

In the development process, the designer and the ship owner were very aware of the safety issues and ergonomic matters, and they explored every avenue in studying all the possible items relating to safety and human behaviors in the operational environment and had technical meetings between the engineer and ship owner (Shimono, 2000). The designers followed the ergonomics principles and user-centered design process. The project started with a solid task analysis to identify what were the functions that speech navigation could be used. They performed the user analysis to identify the operators’ behavior, their needs and design requirements and even the possible vocabulary. The end users were involved in the design process and setting up the design criteria in detail, such as a) the navigator must answer back surely to the spoken command for navigation, b) the system must issue the alert messages such as collision, strike or running on a rock or ashore, deviate or diverge the course and so forth by artificial voice, and c) the system must have a function as the speech memorandum registrations and automatically arranging announcement and so forth. By these considerations, they introduced an adaptable and adjustable speech input system; they limited the vocabulary number, using normal and stereotyped words for navigation, control and management to make the speech input performances sure; they ensured reliable resources of speech technology for maintaining long-term system performance to increase the usability of the system. Introducing the ANSS into ship operation makes team working among the captain, the mate and the engineer change, especially in safety considerations. The socio-technical systems
considerations have been incorporated into the sailing systems design and management. Assessments were conducted from various fields, such as ergonomics and safety, technology, reliability, resources and regulations, and various matters regarding environmental issues.

The success of the integration of speech technology into the navigation system in the Marine ship is mainly due to the factor that the designers put the ergonomics and safety considerations as the first priority and applied the socio-technical systems principle in designing the speech navigation system.

1.3 INTERFACE DESIGN

John Flach once pointed out that "an important foundation for the design of interfaces is a basic theory of human cognition. The assumptions (explicit or implicit) that are made about the basic properties of the human cognitive system play a critical role in shaping designers' intuitions about the most effective ways to communicate with humans and about the most effective ways to support human problem solving and decision making" (1998). This is one of the significant points in the ergonomics, human factors and the socio-technical systems approaches. As outlined in the recent progress of speech applications technology, if the interface design does not accommodate human cognitive demands, fit into the user's application environment, or fulfill social and organizational demands, the product itself will not easily be accepted by the user.

1.3.1 Neuroscience and Cognitive Psychology

A human being is a biological entity. The cognition that happens in the brain has its biological ground. Cognitive psychology tries to discover and predict human cognitive behavior under different environmental conditions while neuroscience tries to explain this behavior from the neural structure and biological function of the brain. In cognitive science, people talk about perception, long term memory and short term memory; about speech and language comprehension; about workload and cognitive limitation; about the process from perceptions to action while neuroscience tries to find evidence on what happens on different perceptions and where such perceptions happen in the brain; how different perceptions connect to each other in the brain; what memory is; and where we have the distinction of short-term memory and long-term memory in the brain; etc. It is very common that people apply different theories to explain certain human cognitive behavior. The best evidence to prove these theories is to find explanation in neuroscience.
The application of cognitive psychology to the system design is what we call "cognitive engineering" and human factors studies. Human factors engineering, socio-technical systems researchers and cognitive sciences have accumulated knowledge regarding efficient and effective user interface design in general. But there is relatively little study regarding the interface design in interactive speech communication systems. There are very limited publications related to human factors and speech interaction design. These limited results of studies will be summarized in different chapters in this book.

1.3.2 Speech Interaction Studies

Basically, the appropriate mode of voice-based man-machine communication contains three different systems: a speech recognition system—a speech understanding and dialogue generation system and a speech synthesis system.

The speech recognition system is the basic part for any voice-based man-machine interaction system. A speech recognition system identifies the words and sentences that have been spoken. The requirement for the recognition accuracy of such a system is different for the different interaction system it is applied. For example, for spoken dialogue systems, to permit for natural communication with humans, the speech recognition system should be able to recognize any spoken utterances in any context, independent of speaker and environmental conditions, with excellent accuracy, reliability and speed.

The speech understanding system and dialogue generation system infer the meaning of the spoken sentences and initiate adequate actions to perform the tasks from which that the system is designed. The difficulty of understanding language depends very much on the structure and the complexity of the sentences and the context in which it has been spoken.

A speech synthesis system generates an acoustic response. The requirement for the acoustic effect of a generated response can be very different depending on the applications as well. For example, for most dialogue systems, it must feel as natural as native speaking to carry dialogue between humans.

The actual state of technology on the speech interfaces may be required to add additional input/output modalities for compensation in order to make the interaction between human and systems seem as natural as possible. This is due to two different reasons: one is because human speech and the automatic speech systems make errors easily; another reason is because human communications are multimodal in nature. As it was demonstrated in the speech sailing system, manual input modality was introduced as the
compromise to the performance limitation of speech recognition systems. Voice portal systems, dialogue through the Internet, are also not the exception of using speech input only. Even though these three principal components of a speech interface are working perfectly, the careful and proper design of the interaction is still important.

1.3.3 Multidisciplinary Aspects

Similar to many other human-system interaction designs, the design of a speech interaction system is multidisciplinary in nature. Knowledge from different disciplines is needed for the design. Sometimes multidisciplinary team collaboration is necessary for a usable product design, and this point has often been ignored. The possible knowledge that is needed for such a design is shown in Figure 1-2.

![Figure 1-2. The multidisciplinary aspects of speech-system interaction design](image)

This figure may not cover all the possible disciplines that may be involved in the speech technology related interface design. If I consider the technical disciplines and design practices in Figure 1-2 as part of system design technologies, then different disciplines under the human factor category would be another important part of the design of the speech interaction system. The knowledge of applied psychology can contribute to the speech technology system since it tells us how to understand human behavior when interacting with the system. The studies on cognition and human factors of speech technology applications together with the socio-technical systems approaches would guide the development of the speech recognition technology in a proper and adequate direction. It provided the guarantee of the usability of the designed system. Very often, people