



OVERVIEW OF COORDINATED RESEARCH IN SPEECH COMMUNICATION IN FRANCE

Jean-Paul HATON*
Director of the GRECO/PRC "Man-Machine Communication"

ABSTRACT

This paper gives an overview of the French national project on automatic speech processing ("GRECO on Speech Communication") funded by CNRS and the Ministry of Research and Technology. We will first concentrate on the role and objectives of the project and then present the six domains which are presently investigated : database of spoken French, database of morphological and lexical aspects of spoken French, acoustic-phonetic decoding of speech, design of a workstation for speech research and development, man-machine dialog using speech, use of prosody in speech recognition.

INTRODUCTION

Research and development activities in the field of automatic speech processing have been relatively important in France for the past twenty years or so. The "Speech Communication" Group of GALF (the French acoustical society) and the working group on "Speech Recognition" of AFCET (the French computer society) have heavily contributed to an informal coordination of the different research groups.

In January 1981 the French national agency for scientific research (CNRS) created a structure, called "GRECO Communication Parlée", with the role of promoting and coordinating public French research activities in the field of speech communication for future applications in automatic speech processing, especially recognition and synthesis. This GRECO (GRECO stands for "Groupement de REcherches COordonnées", i.e. "Group for Research Coordination") is now a part of the national project on Artificial Intelligence and Man-Machine Interface called "PRC Communication Homme-Machine" and sponsored by the Ministry of Research.

We will first give in this paper an overview of this national project stating its goals and means as well as the research groups that participate in its activities. The different projects which are currently supported will then be briefly presented and we will finally draw some conclusions of this 6-year experience of research coordination and give some indications about future activities.

OVERVIEW OF THE PROJECT

General Presentation

The GRECO on Speech Communication created by CNRS was assigned a double objective :

- promote fundamental and applied research in speech communication for application in automatic speech processing (mainly speech recognition and synthesis) by selecting a limited number of research topics and by providing the laboratories with additional financial resources,
- coordinate the activities of the different public groups which carry out research in the field.

Nineteen groups are presently involved in the different activities of the GRECO. They represent an amount of approximately 130 researchers. The GRECO is now part of the national project launched by the Ministry of Research on Artificial

INRIA / CRIN - B.P. 239 - 54506 Vandoeuvre les Nancy cedex, France.

Intelligence and Man-Machine Interaction (called "PRC Communication Homme-Machine"). This large project concerns three related and complementary domains : speech, natural language and vision.

Actions of GRECO

Four types of actions are used by the GRECO :

- support of coordinated projects that necessitate the synergy of various groups with complementary competences (e.g. the design and collection of databases of spoken French, or the implementation of a man-machine dialog system),
- contracts given to laboratories on specific themes,
- working groups which are in charge of investigating new, promising fields that might further give rise to other types of action,
- specialized workshops, with limited attendance, on narrow themes.

Six projects are presently supported :

- design, record and use of a database of spoken French (BDSONS),
- design, collect and use of a database of morphological and lexical aspects of French (BDLEX),
- acoustic analysis and workstation dedicated to speech research and development,
- acoustic-phonetic decoding of speech by machine,
- man-machine dialog using speech.
- use of prosody in Automatic Speech Recognition (just started).

We will now give some indications about these projects.

SCIENTIFIC PROGRAM

The Database for French Spoken Sounds : BDSONS

The construction of a large database for spoken French is one of the major objectives of the GRECO, both for fundamental research on acoustic and phonetic aspects and for the evaluation of automatic recognition and synthesis systems.

The base is presently made up of two types of records :

- records for systems evaluation :
 - . a 2 minute text used as a test pattern to allow users to "tune" their systems ("Rainbow passage" in French),
 - . a list of 100 digits (with a list that can be used for the training pass),
 - . a list of 50 strings of three connected digits (with a training list),
 - . a list of 50 strings of four connected digits,
 - . a list of 50 strings of five connected digits,
 - . a list of 100 numbers between 0 and 99,
 - . a list of 50 spelled words (the letters are pronounced in isolation),
 - . a list of 50 spelled words (the letters are pronounced in a continuous manner),
 - . a list of minimal pairs for rhyme tests, to evaluate the intelligibility of speech synthesis systems.
- records for acoustic-phonetic studies :
 - . a 2 minute text (the test pattern),
 - . a list of the fifteen isolated vowels (with reference words as an aid to pronunciation),
 - . a list of CVCVC strings pronounced 3 times by 4 speakers - 2 male and 2 female speakers - (C is any of the 17 French consonants, V is any of the 15 French vowels),
 - . a list of CVCV pronounced 10 times by the 5 male speakers (V is one of the three vowels /a/, /i/, /u/).

Other records will be added to the base in the future. Up to now 32 speakers have been recorded. The methodology for constituting this database (choice of speakers, recording, etc.) was set up by a working group of the GRECO.

The actual size of the base is 4 gigabytes. Since it is not possible to easily handle such a large database on a classical magnetic support the GRECO has decided to implement the base on video cassettes (Betamax standard) by using a Sony PCM F1 converter and an OROS AI adapter for connection with various mini- and micro-computers. Twelve laboratories have been equipped with this system that makes it very easy to exchange data since the entire database occupies only 18 cassettes. A version of the database will soon be available on CDRoms. An important step in the constitution of BDSOONS consists of segmenting and labelling the acoustic data. This operation is fundamental for the further use of the base by a large number of different people.

The Database for Phonological and Lexical Aspects of French : BDLEX

BDLEX is a complement to BDSOONS, database of spoken sounds. It is intended to provide research groups working in man-machine dialog in natural language with a set of complete linguistic data. Every application in this field necessitates a lexicon, even for very small vocabularies. When the size of the vocabulary reaches some thousand of words it becomes necessary to develop a complete lexicon with its different components : logico-semantic, syntactic, morphological and phonological. That represents a long and tedious job which cannot be directly obtained from usual dictionaries and which constitutes the goal of BDLEX project.

The presently available version, BDLEX-1, is made up of 15 000 roots which correspond to 350 000 words. This base is implemented on Multics and on any PC compatible. The design and realization of this base is a long job that has been achieved thanks to the active cooperation of specialists from various fields : computer scientists, linguists, phonologists, etc. BDLEX represents a fundamental tool for the implementation of man-machine dialog applications both for written and spoken languages.

Acoustic Analysis of Speech

The acoustic data recorded in BDSOONS will be used by the different laboratories for acoustic-phonetic studies about French sounds. In order to introduce some degree of normalization in these studies the GRECO has defined a series of standard programs for the analysis of speech signal. The corresponding Fortran package has been written by some groups of the GRECO and is now available for all laboratories. This package includes :

- four general programs for speech analysis : FFT, LPC, cepstrum and filter banks,
- four pitch analyzers : SIFT, comb, cepstrum and FO-tracking.

A group is now working on the specification of workstations dedicated to speech research and development. This station should include modern means for signal processing and display (VLSI circuits for signal processing, 32-bit general purpose microprocessors, bit-map screen, etc.).

Acoustic-Phonetic Decoding of Speech

Acoustic-Phonetic Decoding includes the various processes after initial signal processing which are involved in the transformation of the acoustic speech wave into a description in terms of discrete phonetic units (phones, phonemes, diphones, syllables). Its role is fundamental in a speech understanding system since the quality of the phonetic decoding will highly affect the overall performance of high-level linguistic processors. As a matter of fact the relatively poor quality of present acoustic-phonetic decoders is certainly the major bottleneck to the development of man-machine dialog systems for advanced applications.

The GRECO has therefore decided to promote coordinated research projects in this area where a good job was already done for the past ten years or so by French groups. Four themes have been selected :

- the definition of multi-context and multi-speaker acoustic cues,
- the formalization of rules and strategies which make it possible to control the transcription of the speech wave into phonetic units,
- the modelization of duration in phonetic units,
- the articulatory interpretation of acoustic events.

These themes imply both fundamental phonetic research and applications to multispeaker speech recognition. Eight contracts have been given to laboratories in order to carry out research in this area.

Man-Machine Dialog by Voice

A multi-disciplinary group (including specialists from artificial intelligence, speech recognition, natural language processing, linguistics, human factors) was set up by the GRECO in 1982 in order to study the various aspects of the design of a man-machine dialog for advanced applications necessitating continuous speech and a natural enough language.

Most of the so-called speech understanding systems developed so far in various countries, including the different systems implemented in the USA during the DARPA Speech Understanding Project, were not developed in a realistic context of man-machine dialog. An important effort of research is still therefore to be done for implementing such systems. In 1985 the GRECO has started with the help of a contract with CNET a study concerning the linguistic and ergonomical aspects of oral man-machine dialog. Two applications were chosen ; both were related to the dialog with an inquiry system accessed by telephone, the first in the case of train schedules and fares in one of Paris's main station, the second in the case of information about graduate studies asked by second-year students at University. In the two cases the speech understanding system was simulated by a human operator. Several hours of dialog recording have been made, labeled and studied during these experiments. The resulting material has been carefully examined both by specialists of human factors and by linguists since we were interested in studying the practical conditions of the dialog and the reactions of the user in a man-machine situation but also in determining the level of language complexity used by the customer.

Use of Prosody in Automatic Speech Recognition

This new project was launched in 1986 in order to study the use of prosodic features (pitch, duration, etc.) at various levels of the understanding process : phonetic, lexical, syntactic and dialog.

CONCLUSION

We have briefly presented in this paper the GRECO on Speech Communication which is a part of the French national project (PRC) on Man-Machine Interface. The role of the GRECO is to promote and coordinate fundamental and applied research in various areas of speech communication for further applications to automatic speech recognition and synthesis. The GRECO was created in 1981 and its influence is obvious in the quality of results obtained and in the mutual knowledge of the different laboratories which participate in the activities. Other projects will certainly be initiated, besides the ones already described in this papers. Particularly, the use of prosody in speech recognition will be considered this year as a new project.