GEMINI

Generic Environment for Multilingual Interactive Natural Interfaces

Summary

Introduction

Recent advances in automatic speech recognition have made it possible to implement automated dialogue services that offer much more than simple yes/no or digit recognition. Modern services require a considerable degree of "naturalness": The user is not forced to use particular words or phrases, neither follow strictly the instructions of the system. At the same time, the size of the vocabulary has been increased, enabling services like directory assistance, or yellow pages, with virtually unlimited vocabulary sizes.

For these reasons, the market is faced with an ever-increasing need for information services the provision of which is based on automatic speech dialogue systems. Information services are provided today by telco's and by independent service providers. At the same time, a number of public bodies and private enterprises, both in and outside the telecommunications markets, realise that they could take advantage of these services to improve their flexibility and significantly decrease the costs of their information services. The public demands enhanced quality in information services in terms of user-friendliness, speed and reliability. These factors are vital for the acceptance of these services. Furthermore, multi-modality ensures that end users will choose the access that best suits them, improving their satisfaction with the service.

GEMINI will exploit experience gained from previous projects and from real-world use of similar systems, to create a generic platform for the development of user-friendly, natural, high quality, intuitive, platform independent and multi-modal interactive interfaces to a wide area of databases employed by information service providers.

Objectives

- User-friendly, natural, interactive dialogue interfaces

- Quick and cost-effective set up

- Adaptability to new applications

- Multi-linguality

- Adaptability to new languages

- Multi-modality

An Application Generation Platform (AGP) containing generic dialogue components available for adaptation to new services and languages will be the main outcome of the GEMINI project. Thus, generation of multilingual and multi-modal interfaces (Speech, Text, Icon Selection, Audio, Video, DTMF tones, etc) will be achieved by incorporating the lexical and semantic relations of the databases contents, reducing the development time and facilitating the system's maintenance and transportability to different applications and languages. Furthermore, the platform will enable a high-degree of personalization (i.e. user modelling, speaker verification, etc).

Two different pilot applications (interfaces) will be developed on the platform, each in three different languages (four languages in total, German, Greek, Spanish and English), in order to test and demonstrate the platform's effectiveness, efficiency and adaptability to new applications and languages.

Expected outcome and innovation perspectives

GEMINI will exploit experience gained from previous projects and from real-world use of automatic speech dialogue systems to create a generic platform for the development of user-friendly, natural, high quality, intuitive, platform independent and multi-modal interactive interfaces to a wide area of databases employed by information service providers.

Specifically, the platform will contain a set of generic dialogue components available for adaptation to new services and languages. A number of sub-modules, already available to the consortium, will be integrated into the platform, while innovative work and research will concentrate on the development of the GEMINI Application Generation Platform (AGP). This work includes the definition of what needs to be described by the system designer, and how it is to be described. Thus, the designer will only have to provide the AGP with the appropriate application-specific information and use the "Language and Application Adaptation Tools" of the platform to adjust the Language and Application Independent Information to the current application, languages and modalities. In this way, generation of multilingual and multi-modal interfaces (Speech, Text, Icon Selection, Audio, Video, DTMF tones, etc) for different types of applications will be achieved faster and in a more cost-effective way, thus reducing the development time and facilitating the system's maintenance and transportability to different applications and languages. Furthermore, the platform will enable a high-degree of personalization (i.e. user modelling, speaker verification, etc).

EG-Banking: this is a Voice portal for carrying out user-friendly, high-quality interactions (information, transfers, payment orders, etc.) for bank customers. e-Banking, phone-Banking and mobile-Banking, are the offered services by this application.

CitizenCare: this is an e-Government platform framework for citizen-to-administration interaction, namely an interactive authority and information guide. All navigating actions and selections will be based on multi-modal (web, speech) dialogue interactions with the user.

Start and duration

1 April 2002, 24 months