DIRHA
Distant-speech Interaction for Robust Home Applications

The DIRHA project addresses the challenge of natural spontaneous speech interaction with distant microphones in a home environment.

The main fields on which research will be conducted, and for which suitable solutions will be identified and embedded in real-time prototypes, are: multichannel acoustic processing, distant speech recognition and understanding, speaker identification/verification, and spoken dialogue management.

The project also aims to investigate the use of a new type of acquisition device consisting of MEMS (Micro Electrical-Mechanical System) digital microphone arrays.

The project addresses four languages: Italian, Greek, Portuguese and German. For comparison purposes, the English language will also be used.

The final prototype will be integrated in automated homes and evaluated by real users.

Targeted application and users

The targeted application includes voice-enabled interaction with appliances and other automatic services available in a household. Although in some cases users could simply try to speak close to the microphone and in a rather controlled way, the expectation is that in the future they would require being able to interact at four-five meters from microphones in a crowded room, with music playing, and other possible active sound sources. For some individuals (e.g. motor impaired), this is a strong immediate requirement, which is the main reason for addressing firstly this category of users under the DIRHA project. To this purpose, a group of possible end-users will be involved from the beginning, in order to define concrete and realistic user requirements. It is foreseen that the most advanced technologies resulting from the project will be integrated in a real-time prototype installed in automated homes, and daily used by the end-users for evaluation purposes.

Objectives and Innovation

One of the most challenging and innovative aspects of the project is the development of a distant speech interaction system, robust to speaker position, even in a noisy and reverberant environment and eventually in a multi-speaker context. Many other projects have recently addressed this concept and tried to realize some early solutions. However, DIRHA will investigate on novel techniques which allow the realization of distant-speech interaction in a multi-room environment and possibly with multiple users.

Among the most relevant innovative aspects, it deserves to be mentioned that acoustic scene analysis will be running in an "always listening" mode (i.e., without the need of any push-to-tak button), with the goal of understanding acoustic/speech activities concurring in the given environment, and eventually delivering speech chunks to the recognition and understanding components. To this end, one needs to realize robust technologies able to tackle unforeseen acoustic environments and noisy conditions. Such goals are new and far beyond the state of the art, not only for an application in the home scenario but also for other domains.

Results of the project

The DIRHA project aims both to make advances at research level in the given scientific fields and to progress at technological level, with the development of a proof-of-concept system which can represent the starting point for a next exploitation action to be addressed by the involved industrial partners.

Research activities will also include the creation of experimental tasks and corpora which will enable initiatives of dissemination and benchmarking at international level.

As for the final prototype, it will run based on microphone devices installed in different rooms in order to monitor selectively acoustic and speech activities observable inside any space of the household. In the targeted scenario, the user can speak from any position in space, i.e. any point in any room of a house given any background noise and acoustic conditions typical of a household, and no matter of where the closest microphones are. A spoken dialogue session can be activated based on a user request, for instance in order to have access to appliances and devices, to services regarding emergency situations, to the media-center (e.g., to search for a given song and play it), or to fill-in and send an SMS message.

Impact

The final objective of the project targets application of automatic speech recognition in four languages with common multimicrophone front-end, spoken dialogue management, and user interface. This will have a relevant impact in terms of synergetic approach to the development of spoken language interaction systems and to the immediate evidence of a possibly easy portability to other languages. The project would also represent a milestone for developers and integrators of home automation systems, since the targeted prototype can be a first proof-of-concept realization in a real world context, based on concrete and realistic user requirements and operational constraints.

The DIRHA consortium aims to examine the impact of its novel technologies primarily with collaborative users. In other words, the DIRHA system will be conceived for subjects who have, in principle, no difficulty in understanding the way to access the system in order to obtain the highest satisfaction (e.g., based on a high completion rate in the proposed tasks) and who have a very good attitude towards this experimentation. Once the basic technology has been established and evaluated as reliable, other categories of users (e.g., elderly people) may be addressed in future projects.

Another impact of the project regards the portability of the foreseen solutions to other possible domains. In fact, the DIRHA approach and the resulting technologies could eventually be applied to several application contexts characterized by noisy environment and by the need of talking far from the microphone as, for instance, robotics, surveillance, telepresence, gaming, industry sector and manufactory.