Publishable Final Activity Report

Project reference: FP6-034624
Project acronym: DICIT
Project title: Distant-talking Interfaces for Control of Interactive TV
Instrument: Specific Targeted REsearch or Innovation Project
Thematic priority: FP6 - Priority 2 – Information Society Technologies
Start date of project: 01st October 2006 Duration: 36 months
Date of submission: 29th December 2009 (revised in February 2010)
Coordinating organization: Fondazione Bruno Kessler (FBK) – Center for Information Technology - irst, Via Sommarive 18, I-38123 Povo, Trento, Italy
Project Manager: Maurizio Omologo
Tel. +0461-314563 Fax: +0461-314591 e-mail: omologo@fbk.eu

Other Contractors: Amuser, Torino, Italy
Elektrobit Automotive GmbH, Erlangen, Germany
Fracarro Radioindustrie S.p.A., Castelfranco Veneto, Italy
FAU (Friedrich-Alexander Universitaet) Erlangen-Nuernberg, Erlangen, Germany
IBM Ceska Republica, Praha, Czech Republic
Alpikom, Trento, Italy
Contents

Publishable Final Activity Report ........................................................................................................... 1

1. Project Objectives and Expected Results .......................................................................................... 4
2. Methodologies, Approaches, Work Performed and End Results ...................................................... 6
3. Dissemination activities .................................................................................................................. 11
4. State-of-the-art, achievements, and possible impact of the project ............................................... 12
5. Publishable results of the final plan for using and disseminating the knowledge ...................... 14
Coordinator

FONDAZIONE BRUNO KESSLER

Partners

Website address: http://dicit.fbk.eu
1. Project Objectives and Expected Results

The target of the DICIT project was to progress on technologies for multi-microphone signal processing, distant-talking speech recognition, and spoken dialogue management, particularly when applied to multi-modal human-machine interaction.

The main reference scenario (see Figure 1) comprised a multi-modal user-friendly interface that allows to have access to a virtual smart assistant enabling the interaction with TV-related digital devices and infotainment services. The user can speak in a comfortable way, not encumbered by any hand-held or head-mounted microphone, at a distance of up to some meters from a set of microphones, complementary to usage of the remote control. The environment can be a living room equipped with digital TV, Hi-Fi audio devices, etc., and populated by a group of people (e.g., family members). Any of the given subjects, one at a time, can have access to the system with the goal of querying information about an upcoming program, scheduling its recording, and so on.

Figure 1: Scenario of multi-modal control of TV and related devices in DICIT.

One of the most challenging aspects of the given scenario is that, during the interaction, the TV and related devices are producing their own sound, for instance by means of a set of loudspeakers. This requires the integration of the speech acquisition
module with an adequate Multi-channel Echo Cancelling system. The objective is to obtain an input signal of sufficient quality to perform satisfactory Automatic Speech Recognition (ASR). A challenge here is the operation in a real-life noisy and reverberant environment with multiple acoustic sources. Source Localization, Source Separation and Adaptive Spatial Filtering by beamforming are other techniques of choice which needed to be perfected for the given task.

Another significant objective was the implementation of continuous listening by the system, which required the development of a smart Acoustic Scene Analysis component, which embeds detection, localization, classification and interpretation of all the possible acoustic events, including non-stationary disturbances. To achieve ASR robustness, it was also necessary to exploit both an advanced front-end processing and a robust acoustic and language modeling: Adaptation was a key issue, to make the system able to learn about changes in the environment and adjust to the general context of interaction. In this regard, Speaker Identification and Verification, as well as system adaptation to the user, although technically very challenging, represented other research issues to explore at a preliminary level. Natural Language Understanding (NLU), Mixed-initiative Dialogue Management, and proper Response Generation represented other fundamental components for the realization of the proposed smart interface.

A crucial aspect was the design of a suitable Architecture for the integration of all the modules developed by different partners and the realization of intermediate and final prototypes. Due to real-time requirements and known limitations of embedded platforms typically used for SetTopBox (STB)-class devices, a careful analysis was necessary to find the optimum compromise between feasibility, portability, and effectiveness, with the eventual intention of incorporating the resulting technological components into commercial products.

The main expected result was a prototype, resulting from the best combination of the technologies developed during the project, comprising a real STB platform, a remote control device, a microphone network and a multichannel loudspeaker system for voice input and general audio output. The targeted prototype had to handle a multi-modal spoken dialogue in three languages, i.e., Italian, English and German. The language of the interaction, although constrained by a limited size of the (dynamically updated) vocabulary, was to be adapted to provide the user a perception of a natural interaction with an intelligent assistant. Moreover, it was desirable that the targeted system should react immediately to user requests even when the audio output system is active, i.e., also implementing a barge-in functionality.

Finally, in order to show the portability of DICIT technologies to other domains, in particular the multi-microphone front-end to other application areas such as smart-
home, surveillance, security, video-conferencing, surgery room, games, etc., the home surveillance was considered as a second scenario to explore.

The realization of an anti-intrusion prototype as well as of the intermediate and final STB-based interactive-TV prototypes represented the three main milestones of the project.

2. Methodologies, Approaches, Work Performed and End Results

Year 1
At the beginning of the project a Market Study was conducted both to explore existing solutions, and to understand user needs and requirements, through specific focus groups. The results of that study provided very useful information about the possible trend of this market, about the expectations of the users and, consequently, about the system to design. A set of WOZ (Wizard-of-OZ) experiments was also conducted to better understand the way people would use the targeted system and, consequently, to design the first STB prototype user interface. Data were also collected for the design of the spoken dialogue management component as well as of the multi-microphone front-end.

During the first year, the consortium was active in defining the prototype architectures. In particular, the IBM - CIMA (Conversational Interaction Management Architecture) framework was adopted to run SCXML applications based on the DICIT dialog specifications.

As to the multi-microphone front-end, a stereo Acoustic Echo Cancellation module was adapted to work jointly with beamforming in an effective way given the first STB prototype requirements. Advances in speaker localization techniques were achieved, in particular with the design of a new GCF(Global Coherence Field)-based solution to handle potential multiple speaker conditions. A speech activity detector was also developed for real-time robust speech recognition with distant-talking interaction.

Distant-talking ASR was addressed studying the impact of front-end processing on system performance. To train acoustic models, contaminated databases were produced for the given environmental acoustics and languages. Acoustic Model adaptation was shown to improve the expected ASR accuracy for the first STB prototype. Additionally, the IBM-Embedded ViaVoice ASR engine was successfully integrated into the CIMA package which then formed the basis for the first prototype.

Several other key components were implemented, including NLU engine, Dialog Manager, TTS engines, and GUI output. The multimodal dialog flow was designed including prompts, TV layouts, and usage of a User Profile.
During the second year, activities progressed at FAU and FBK realizing a multi-microphone front-end that was able to react in real-time to distant-talking speech input by: localizing the active speaker, beamforming towards the corresponding direction, performing acoustic echo cancellation to remove contributions from the loudspeakers, applying an accurate speech activity detection that rejects possible undesired disturbances, and eventually providing a speech chunk as input to the recognizer.

A robust recognition engine was obtained, based on an acoustic modeling trained with regard to the specific conditions of the distant-talking interactive TV scenario, which includes reverberation, background noise, effects of beamforming and residual of echo cancellation. Language modeling was set up for the English Language collecting more than 15000 written utterances from a panel of potential users; then the Italian and German versions of the Language Model were derived by translating and integrating the English one.
The first STB prototype (see Figure 2) included all the functionalities as planned in the project workplan. It is worth noting that the multi-modal dialogue application running on the given CIMA framework was developed by using Elektrobit tools, namely EBGUIDE Studio and EB GUIDE. Moreover, the system included a real STB device, which was designed and developed by Fracarro according to the given specifications. The prototype supported English, German, and Italian languages, and it was installed in the premises of two industrial partners, i.e., Amuser and Elektrobit. An evaluation campaign was conducted with real users, for which specific evaluation methodologies, criteria and tools were defined and created.

The distant-talking speaker identification task was also explored, with a progress that allowed developing another prototype including a related functionality in the given real-time multi-microphone front-end; the system implemented a simplified TV command-and-control task able to operate under more challenging noisy conditions.

Acoustic event detection components were then developed for the anti-intrusion surveillance prototype, which consisted of two main components, a PC platform and an Alarm Panel. The prototype was installed at FBK, and the related evaluation campaign showed a satisfactory accuracy both in detecting an intrusion event and in discriminating various events occurring either inside or outside the environment, the latter one being a very important feature for a specific real-world application of the technology.

Year 3
As an intrinsic step of the cyclic approach characterizing the project workplan, the evaluation of the first STB prototype had provided the consortium with many indications on which aspects to emphasize in the design of the final prototype, which had to represent the synthesis of efforts and achievements attained during the project life. The main limitations of the first prototype were due to an insufficient recognition performance and to a relatively large delay in the system response. A first action was devoted to address and enhance these critical aspects, which led to releasing a set of new components for all the layers of the prototype, whose basic architecture however was not altered with respect to the first one.

The final STB prototype was then realized (see Figure 3), including a more advanced multi-microphone front-end able to acquire selectively speakers and to distinguish between the main user, located within a designated area, and other possible interfering speakers localized outside of it. The front-end was characterized by a further improved performance of acoustic echo cancellation and beamforming modules as well as by a more robust smart speech filtering that accurately analyzes and classifies speech chunks, before sending them to the ASR. ASR and NLU performance were also improved taking into account more accurately the effects of the environment and of the DICIT front-end pre-processing, and aiming at an
improved understanding accuracy with vocabularies and languages typically adopted under real usage of the prototype.

**Figure 3:** Hardware architecture and components of the STB prototype.

Improved graphical user interface and dialogue flow (both with speech and remote control) were implemented based on the experience acquired during the first prototype evaluation. Based on that, a simple user profiling and some help functionalities were also added. EPG setting was based on downloading the related information from Internet, which provides up-to-date information about current programs on the satellite (see Figure 4). A new STB device was produced, with more efficient and effective hardware and firmware modules which enable fast channel switching, Common Interface for scrambled programs, faster communication with Dialog Manager, and new layouts for the On-Screen-Display (OSD) interface.

A very important activity of the third year was the evaluation of the final STB prototype. In order to ensure a comprehensive analysis on the effectiveness of the proposed technologies and solutions, the final evaluation campaign involved all the partners’ sites (see Figure 5) where prototypes were installed. Due to a possible variability in the operating noisy and acoustic conditions as well as in the behaviour
of hardware and devices, which in some cases could differ from one site to another, a calibration procedure was defined. Thanks to very accurate hardware and software settings, it ensured a coherent performance evaluation across different installations.

Figure 4: Example of EPG (Electronic Program Guide) screen in the final English STB prototype. The user can input a natural language query to select a title, a channel, or to perform a filtering for a more selective search.

Methodologies, criteria, and tools for the final STB prototype evaluation were defined, with the purpose of supporting both performance evaluation for several internal modules of the prototype and usability evaluation of the dialogue. The prototype was tested with 172 subjects, in the three languages, and at seven partner sites. The results showed a clear improvement concerning dialogue completion time and speech recognition accuracy, which were two weaknesses of the first prototype. The subjective feedback by the participants was strongly positive and the objective metrics yielded satisfactory results. As an important indicator of this experimental evidence, the average task completion rate increased from 50% to 85% from the first to the final prototype.

Besides the activities related to the final prototype development and evaluation, the consortium progressed in other technical fields. A new speaker verification system based on phoneme class segmentation was developed, which is optimized for short utterances and for the possible effects of the residual echo in the input signal. A novel combination of Stereo Acoustic Echo Cancellation (2C-AEC) and Blind Source
Separation (BSS) algorithms was also investigated, and a real-time system combining 2C-AEC with determined or underdetermined BSS has been set up. Finally, a comparative analysis of a classical GCF-based source localization and the BSS inherent source localization was conducted.

**Figure 5**: Subjects using DICIT prototypes at FBK (left panel) and FAU (right panel).

### 3. Dissemination activities

Different dissemination activities were conducted during and at the end of the project. In the following a short list is reported including the most relevant contexts in which these actions were accomplished.

An official and publicly available **Web site** was established (see http://dicit.fbk.eu). Most of the information about the project achievements are downloadable, namely, public deliverables, papers, video-clips showing use of various prototypes, references to presentations of the project at conferences, workshops, and fairs, instructions on how having access to resources (e.g., WOZ data) created during the project.

A DICIT event was held in Trento, as satellite of IEEE-HSCMA (Hands-free Speech Communication and Microphone Arrays) 2008 Workshop (see http://hscma2008.fbk.eu), which hosted about 100 researchers.

Invited talks, distinguished lectures, and seminars describing the objectives and the main achievements, were given in events worldwide, as for instance at ACOUSTICS 2008, IFA 2008, JEITA 2008, and EUSIPCO 2009. The scientific production concerned the publication and presentation of papers at major conferences based on peer review, as ICASSP, INTERSPEECH, IWAENC, etc.
It is worth noting that DICIT prototypes were also presented in two very important appointments, i.e., at ICT 2008 in Lyon and at IFA 2009 in Berlin. In both cases, the prototype was running in an open-space under very challenging noisy conditions, which showed its robustness (see Figure 6).

Interviews were given to newspapers, radio, and TV. In particular, a service of Euronews was dedicated to the presentation of DICIT at IFA 2009.

During the project, the consortium met AMIDA, LUNA and VISNET-II EC project consortia with the objective of discussing on methodologies and tools for evaluation of some of the addressed technologies, as for instance multi-microphone space-time audio processing, spoken language understanding, and spoken dialogue systems.

Figure 6: Presentations of DICIT prototypes at ICT 2008 (left panel) and at IFA 2009 (right panel).

Finally, it deserves mentioning that a Restricted Focus Group (RFG) of industries was set up during the first year. The purpose of this action was to have an external advisory team providing information about how the project was technically sound and how it might be redirected to better focus on the most relevant issues. The initial activities with RFG were very fruitful, and with some of the members exploitation issues were addressed.

4. State-of-the-art, achievements, and possible impact of the project

In October 2006, at the beginning of the DICIT project, there were not many examples of frameworks under which the combination of multi-microphone processing, distant-talking speech recognition and understanding technologies had
been investigated and successfully applied. In particular, in a scenario as the DICIT
one, performance was limited by reasons that can be summarized again as follows:

i. drastic performance degradation of speech recognition for distant talkers even
in ordinary home environments, due to reverberation, high noise levels, strong
interferers, and multiple speakers;

ii. variabilities in the characteristics of the acoustic sources that should be
accurately classified and interpreted for a better comprehension of the given scene;

iii. performance of standard acoustic echo cancellation algorithms, insufficient
to allow barge-in in multichannel reproduction situations;

iv. lack of adaptation and customization of the conversational speech interface to
the needs and demands of its potential users;

v. limited degree of integration of voice and other modalities.

At that time, some devices were available in the market which tried to integrate
remote control and speech recognition to control a STB platform. However, the use
of those devices was inconvenient since the set of admitted commands was quite
restricted, the system was trained to work with a single specific user, and the
maximum user-microphone distance to obtain acceptable recognition performance
was rather short (e.g., not more than one meter). With a limited diffusion of those
devices, the market was still reflecting the fragility of those solutions.

During the recent three years, the situation has not changed substantially. The
application of speech recognition in the home automation field has been extended
significantly, also thanks to advances of the basic technology that can be found in off-
the-shelf software packages. However, these products are still supporting single
commands or scripted interactions, they often operate based on speaker dependence,
and they require a close-talking voice input (see products as Accenda, HAL,
OneVoice, etc.). In general, flexible solutions based on distant-talking input as well as
on natural language and multi-modal spoken dialogue are still at prototypical and
demonstration stage.

Overall, the DICIT project has succeeded in progressing in most of the
technological fields related to the targeted scenario. The final STB prototype results
from a balanced combination of quite well working components, integrated one to the
other in an effective way at the various system layers in order to obtain satisfactory
understanding accuracy, system promptness, a user-friendly interface, and overall a
perception of smartness at user level. In this type of systems just a bottleneck as, for
instance, the delay in the response of few components characterizing the first
prototype, can compromise definitely the impact of the technology on the user.
Moreover, improvements in the performance of a single component does not
necessarily grant a perceived better system performance at user level.
While the project achievements can be considered substantial, the current prototype should be primarily seen as a starting point for next research activities aiming to improve performance of the given family of technologies. A common opinion, in particular provided by European industries which were contacted or involved in the project, is that at this moment the technology is not mature to tackle the development of a DICIT-like product for the consumer market. It is felt that at least three-five years are still necessary, due to the high complexity of the current solution, to the possible cost that would have a commercialized platform, to all the variabilities in a daily usage of these systems yet not investigated enough, and to many other possible reasons which would determine a probable lack of acceptance at end-user level. An increased system robustness and self-adaptation capability are probably the most urgent issues to address together with aspects related for instance to portability, miniaturization, and localization on other languages. On the other hand, requests for collaboration on the very same interactive TV application from globally leading companies (Sony, Samsung) underline that the project addressed the right problem at the right time.

Moreover, several exploitation actions in a shorter-term seem to be feasible for less complex applications, as for instance to introduce the distant-talking interaction in rather controlled command-and-control tasks for a home automation directed to disabled users. Some knowledge and tools can then be exploited in the development of more advanced voice-enabled systems in the automotive field. Multi-microphone front-end processing seems to be suitable both for smart features to introduce in video-conferencing systems and to start activities towards the realization of anti-intrusion systems for domestic environment. Other application fields can benefit from the DICIT achievements, as in the cases of gaming, surgery rooms, etc., and in general wherever distant-talking interaction can be fruitfully combined with speech recognition and understanding of simple queries given under rather controlled environmental conditions.

5. Publishable results of the final plan for using and disseminating the knowledge

Several potentially exploitable results were identified at the end of the DICIT project, as reported in the following. Note that for most of them no information can be provided at this moment as far as patent registrations are concerned.

In the following, for each technological component a description of the result so far achieved, of possible applications, of the actual stage of development and in case of related IPR is given.
Acoustic Source Localization

- **Result Description**
The localization technique based on Generalized Cross Correlation – PHAse Transform (GCC-PHAT) and Global Coherence Field (GCF) was adjusted to operate in the interactive TV scenario with multiple speakers and possible disturbances, with signals acquired by a linear microphone array and the necessity of ignoring the contributions of the TV loudspeakers. In the surveillance scenario instead, a distributed network of microphone pairs was adopted.

- **Possible applications**
Home automation, surveillance, audio-video teleconference, entertainment

- **Stage of development**
Prototype running in real-time on PC

- **Intellectual property rights**
Based on patented technology of FBK

- **Contact details**
FBK

Beamforming

- **Result Description**
Beamforming for sound capture of single and multiple sources by a linear microphone array with nonuniform sensor spacing was adjusted to the interactive TV scenario. Generic beamforming design methods were optimized to enhance speech recognition rates and to interact efficiently with acoustic echo cancellation. Multiple beams are implemented for capturing several sources.

- **Possible applications**
Home automation, surveillance, audio-video teleconference, entertainment

- **Stage of development**
Prototype running in real-time on PC

- **Intellectual property rights**
Based on technology previously published by FAU

- **Contact details**
FAU

Multichannel Acoustic Echo Cancelling

- **Result Description**
A known multichannel acoustic echo cancellation (MCAEC) algorithm was customized and optimized to interact with adaptive beamforming methods as used for DICIT. The known MCAEC concept has been extended to the case when multiple local sources require multiple beams, and, therefore, multiple MCAEC structures need to be realized efficiently. The crucial issue of mutual decorrelation of reproduction channels is solved by a novel psychoacoustically motivated time-variant phase modulation scheme.

- **Possible applications**
Home automation, surveillance, audio-video teleconference, entertainment
- **Stage of development**
  MCAEC implemented for multiple beam running in real-time on PC
- **Intellectual property rights**
  Based on technology previously published by FAU. The psychoacoustically motivated time-variant phase modulation scheme is patented jointly by FAU and Fraunhofer Gesellschaft.
- **Contact details**
  FAU

**Blind Source Separation**

- **Result Description**
  The TRINICON-based BSS algorithms are currently adapted to the DICIT scenario by constraining the angular range of the direction of arrival for the desired source and by extracting sources also in noisy and underdetermined scenarios, i.e., where more sources are present than sensors used by BSS. At the same time, TRINICON-based BSS is able to localize multiple sources.
- **Possible applications**
  Home automation, surveillance, audio-video teleconference, entertainment, hearing aids.
- **Stage of development**
  The implementation including it inside the DICIT prototype is ongoing work
- **Intellectual property rights**
  Based on technology previously published by FAU.
- **Contact details**
  FAU

**Acoustic Event Detection**

- **Result Description**
  Different features were exploited to detect acoustic events according to the operative scenario. In a quiet living room signal dynamics is sufficient, while considering also spatial coherence of the sources makes the detection more robust in case of reverberation or diffused background noise. For the anti-intrusion system an additional feature related to spectral variations was introduced to improve the detection sensitivity.
- **Possible applications**
  Surveillance, front-end for hands-free speech interaction
- **Stage of development**
  Prototype running in real-time on PC
- **Intellectual property rights**
  Based on technology previously developed by FBK
- **Contact details**
  FBK

**Acoustic Event Classification**

- **Result Description**
  A new classification scheme was adopted, combining the scores provided by subsystems
based on Gaussian Mixture Models and on Support Vector Machine. The combined classifier produces improved results in the case of acoustic events acquired by distant microphones in noisy and reverberant environment.

- **Possible applications**  
  Surveillance, front-end for hands-free speech interaction

- **Stage of development**  
  Prototype running on PC.

- **Intellectual property rights**  
  Based on technology previously developed by FBK

- **Contact details**  
  FBK

**Speaker ID and verification**

- **Result Description**  
  A combination of four subsystems differing in the classification technique and in the trained acoustic models was implemented. Two classifiers were adopted: GMM and Support Vector Machine (SVM). Experimental results show that the new system allows a considerable performance improvement, especially under the critical position mismatch conditions, when trained models and test utterances are referred to different speaker locations.

- **Possible applications**  
  User profiling, authentication, access limitation and safety in distant-talking systems

- **Stage of development**  
  Prototype running on PC

- **Intellectual property rights**  
  Based on technology previously developed by FBK

- **Contact details**  
  FBK

**Acoustic Modeling and front-end processing for distant-talking ASR**

- **Result Description**  
  Acoustic Models are produced considering the effects of room acoustics when the talker is at some meters distance from the array. The whole acquisition and preprocessing chain was taken into account to create models well-matched with the operative conditions of the DICIT prototype. The data contamination approach was exploited; acoustic data directly acquired on the field allowed a more matched adaptation and hence better recognition performance.

- **Possible applications**  
  Front-end for hands-free speech interaction systems

- **Stage of development**  
  Off-line data processing and model adaptation

- **Intellectual property rights**  
  Based on technology previously developed by FBK and IBM

- **Contact details**  
  FBK, IBM
Language Modeling and NLU for the interactive TV scenario

- **Result Description**
  Natural Language Understanding (NLU) based on Statistical Language Models for speech recognition and interpretation enable users to make their requests in a flexible manner without being restricted to specific key-phrases. This is particularly important in the interactive TV scenario as there are many functions – especially when navigating the electronic program guide. Pairs of language and interpretation models were developed for key dialog states along with grammars for variable content such as channel names, genres, numbers, dates, times etc. Models were developed for English, Italian and German.

- **Possible applications**
  Language model based interaction is being routinely employed in many speech applications such as call-routing, music selection, address specification for navigation etc.

- **Stage of development**
  Language models were developed for both the first and final STB Prototype.

- **Intellectual property rights**
  Based on IBM’s NLU technology.

- **Contact details**
  IBM

**CIMA**

- **Result Description**
  CIMA (Conversational Interaction Management Architecture) provides a flexible architecture for multi-modal Dialog Management and for component integration. CIMA is the central integration point for all components running on PC2. The CIMA Dialog Application (logic and prompts) which defines the entire dialog flow for the prototype was developed using the EB Guide Studio modeling environment from Elektrobit. In addition CIMA spokes (interface functions) were developed to integrate the various components of the system – namely the front-end audio processing, speech recognition and text to speech engines, STB, remote control and Electronic Program Guide database.

- **Possible applications**
  Dialog management is deployed in speech applications such as for transaction processing, music selection management, points of interest and address input management etc.

- **Stage of development**
  CIMA application has been developed for first STB Prototype and updated for the final one.

- **Intellectual property rights**
  CIMA is IBM’s Dialog Management technology.

- **Contact details**
  IBM

**EB Guide**

- **Result Description**
  EB GUIDE provides a platform for the modeling, mock-up, and testing of multimodal, speech-enabled user interfaces. EB GUIDE is the central tool to realize the design the
DICIT user interface based on the dialogue specifications. Within DICIT an export filter for GUIDE was developed that provides a dialogue definition which can be immediately used by the dialogue manager CIMA. The export function also supports the NLU capabilities of CIMA.

**Possible Applications**
The connection between a specification/modeling tool and the dialogue manager is one important step towards a seamless development environment for NLU-capable, multimodal dialogue systems. Currently, a complete tool chain for NLU is not available on an industry-scale. However, there is a demand for such tool in various business areas, e.g., the automotive business.

**Stage of development**
EB GUIDE itself is fully developed and commercially available. The extension which connects EB GUIDE to CIMA is usable and stable. Further improvements, in particular with respect to the handling of language models are necessary.

**Intellectual property rights**
EB GUIDE is EB’s modeling technology.

**Contact details**
EB

---

**Evaluation Framework**

**Result Description**
The evaluation framework provides tools for the semi-automated evaluation of dialogue systems based on user interaction data. The system supports data acquisition, visualization, analysis, and annotation such that evaluation can be made much more efficient.

**Possible Applications**
The EvaluationSimulator tool, which is part of the framework, does not depend on a specific domain and is not confined to DICIT. Other application areas, e.g., for the evaluation of dialogues systems in cars, are possible.

**Stage of development**
A working prototype of the framework usable for DICIT exists. Further effort has to be invested to generalize the approach and make it available for other application domains.

**Intellectual property rights**
The evaluation framework is EB’s IP.

**Contact details**
EB

---

**Speech Corpora**

**Result Description**
Various corpora were collected through WOZ experiments, for multi-microphone front-end development and test, for speech recognition and language modelling training purposes. The FBK portion of acoustic WOZ data is freely available upon request (see DICIT website).

**Possible Applications**
Research on acoustic scene analysis based on multi-microphone front-end.
• **Stage of development**  
  Data are available. Annotation is being refined.

• **Intellectual property rights**  
  FBK is the owner.

• **Contact details**  
  FBK