1. Publishable Summary

Project Objectives

The project goal is to design and build mobile applications that approach human performance in conversational interaction, specifically in terms of the interactional skills needed to do so, such as recognising and generating conversational speech incrementally in real-time, adapting to new information and reacting to new communicative situations. All of these skills will be learned or adapted using real data, and used to build systems for voice-enabled interactive search and information provision applications. Current search engines work well only if the user has a single search goal and does not have multiple trade-offs to explore. For example, standard search works well if you want to know the phone number of a specific business but poorly if you are looking for a house with several different search criteria of varying importance, e.g. number of bedrooms versus bathrooms versus price etc. The latter requires the user to collaborate conversationally over several turns.

To realise this vision, PARLANCE aims to make advances in robust, incremental human-machine interaction, personalisation, and adaptation. For PARLANCE System 1, the goal is to create a system that is *incremental* and *hyper-local*. Traditionally, the unit of speech has been a whole utterance with strict, rigid turn-taking determined by a voice-activity detector. By creating a system that is incremental, we hope to improve the experience for the user and create more natural interactions.

PARLANCE System 2 will be *personalised* and *adaptive*. PARLANCE will target the learning of interactional skills for coping with unforeseen concepts thus moving beyond the current focus on limited application domains that use fixed, static databases. We will employ dynamic User Models containing information such as individual's current location, search history, social context, preferences, and a language profile. PARLANCE interactive search will be hyper-local, that is: it will use the user's location and entities in the very close proximity to not only provide more relevant searches but also to improve the accuracy of the understanding of the system. This type of search will be extremely useful in hands-busy and eyes-busy situations, such as driving and walking.

Key to PARLANCE is that all information used by the system is represented probabilistically via dynamic Bayesian Networks. Each user input and search provides evidence that is used to update the probability models and all system responses depend on a stochastic mapping from the network probabilities to actions. Changing local contexts including geographic location will change the conditioning on critical variables, allowing optimisation of strategies even when the information available contains noise. This approach is fundamentally different to previous approaches based on logic and symbolic reasoning, which are often extremely fragile in the face of high error rates and uncertain user behaviour.

Finally, this project will be facilitated by the recent development of cloud-based data collection and evaluation platforms (e.g. Mechanical Turk). These recent advances are particularly timely for PARLANCE, allowing us to more easily and cost effectively collect data. Such data collection and evaluations will provide a new shared data-resource not only for PARLANCE but for all researchers working on next-generation natural speech applications for whom the data will be freely available.

Work performed since the project start:

- System Architecture design and requirements analysis (D6.1);
- Evaluation strategy and metrics for System 1 (D6.2);
- Micro-turn architecture to enable incremental speech recognition and natural language understanding: design of architecture and initial prototype (D1.1);
- Initial work on extending the POMDP Interaction Manager and simulator to handle micro-turn dialogues and learn appropriate strategies (to be written up in D2.1 month 16);
- New statistical approaches to incremental Natural Language Understanding (D1.1);
- Investigation into methods for open information extraction that combine shallow syntactic and semantic features (to be written up in D4.1 month 16);
- Initial architecture of the adaptive Knowledge Base component, including the Ontology Manager and the User Model (D6.1);
- An interactive hyper-local search API (to be written up in D5.2); and
- Geographically aware content discovery (D5.1).

Main results to date

As planned we have an initial architecture design for both System 1 and the full System 2 (see Figure 1 and D6.1). This is a highly modular system that enables incremental dialogue as well as adaptivity, personalisation and hyper-local search.

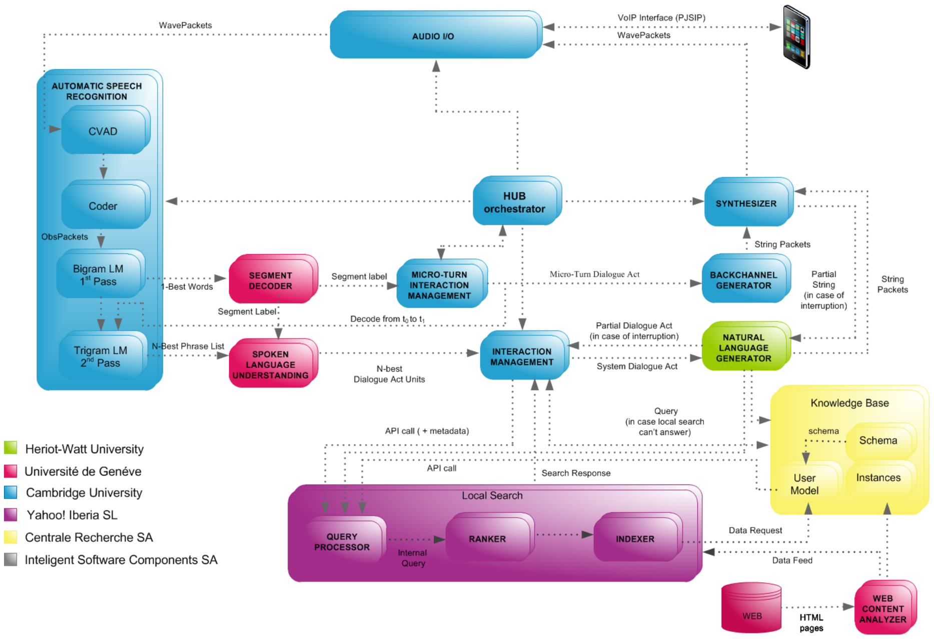

Figure 1: Overview of the PARLANCE System 2 Architecture

An initial prototype of part of this system was completed in Year 1, namely the micro-turn speech subsystem where the micro-turn controller uses the incremental ASR output to decide whether to signal a turn change or not. Currently, the user turn is signalled when a non-filler word is found in the incremental output. While in user turn state, if the incremental output offers limited information, the system outputs a short back-channel and stays in user turn state. This involved developing fundamental software components necessary to support incremental ASR by performing two passes (see Figure 2), SLU and micro-turn Interaction Management, as well as a separate Voice Activity Detection module (see D1.1). This has been implemented for English and work has started on a Mandarin variant.

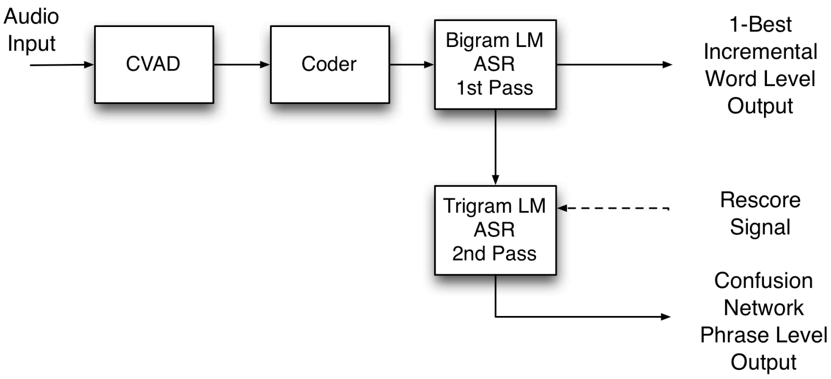

Figure 2. Micro-turn ASR Configuration

The PARLANCE INCremental Dialogue Act (PINC) scheme (see D6.1) was developed, which decomposes traditional dialogue acts into smaller units or *primitive* dialogue acts that can be recognised incrementally from partial user utterances. Human-human data from the EC FP7 project SpaceBook (270019, FP7/2011-16, see http://www.SpaceBook-project.eu/ and (Janarthanam et al. 2012)) was shared with the PARLANCE consortium and has been partially labelled with this PINC scheme. These data will in turn be shared with the scientific community.

With regard to Natural Language Generation and spoken output, an implementation of a prototype system for incremental dialogue was developed that uses reinforcement learning in flat and hierarchical settings. This prototype extends the state-of-the-art in several ways, most significantly through the application of statistical techniques for the automatic optimisation of system behaviour and by being more context-aware and human-oriented than current incremental surface realisers. A crowd-sourcing evaluation showed this method to have more positive feedback when compared against a number of baselines. Regarding speech synthesis, an in-domain corpus was developed for expressive speech synthesis targeted to the application, and an initial expressive HMM voice was trained.

The architecture of the knowledge base including the Ontology Manager and User Model were defined and a novel approach to ontology module population was developed using dependency analysis and a web search engine Pattern detection techniques for text-based information retrieval were developed and extensive experiments were carried out to enrich modular ontologies automatically with new attributes and values by mining the Web. Using this method, the accuracy of the discovered attribute values for 15 attribute types is between 72-87%, using a corpus including 50 snippets for each attribute type. Finally, an improved method for bootstrapping relations from a small set of seed tuples was developed, which potentially outperforms state-of-the-art iterative bootstrapping methods, scales well to large corpora, and requires only minimal linguistic analysis.

As well as creating a search service for the domain as an interactive hyper-local search API which was done ahead of schedule, there have been key developments in hyperlocal search including methods for discovering points of interest in unstructured text; assigning geographic coordinates to a point of interest within a median accuracy of 0.96 kilometres; and categorising points of interest with 70% accuracy.

By the end of Year 1, the consortium partners started the integration of System 1, an evaluation plan to test this system using crowdsourcing was proposed and a number of evaluation metrics and methods provided.

Expected final results

The PARLANCE project is expected to produce the next generation in data-driven spoken dialogue systems that are incremental, adaptive, and personalised, and that use a dynamic knowledge base populated from hyperlocal search. The PARLANCE project will create this dynamic, incremental system in English and Mandarin; finally, in order to demonstrate portability to new languages a system in Spanish will also be created with a subset of functionality of the main PARLANCE system.

Potential impact and use

PARLANCE has the potential to become embedded in everyday life for millions of people. As the amount of information on the web increases at a huge rate, in the near future, interactive, personalised search will be the only way to digest and explain this information to the overwhelmed user. Natural, interactive search will provide vital conversational information access to many sectors of the population.

The PARLANCE programme targets significant impact in the scientific community and in commercial applications of speech technology. Through its work on incrementality, data-driven techniques and adaptivity, it will enhance the naturalness of spoken dialogue systems, making them more likely to be adopted in everyday life. In addition, PARLANCE will move away from the traditional method of creating systems, reducing development costs and time-to-market, and stimulating innovation and expanding markets. PARLANCE is also expected to have a significant impact on the European competitive position in a multilingual digital market by providing improved services to citizens.

Figure 2: Project Logo

Project website: http://www.PARLANCE-project.eu

Contact: Helen Hastie (Coordinator): h.hastie@hw.ac.uk