1 Publishable summary

1.1 Summary of Project Description and Objectives

In the last two decades, interactive spoken dialogue systems (SDS) have moved from research prototypes to real-life commercial applications. SDSs are now pervasive and represent a business with yearly revenue of over 1B Euros worldwide. Still, one major roadblock in commercial SDS prototyping is that the demand of significant effort and expertise for the *development* of new services, the *enhancement* of the performance of deployed services and the *customization* of services to specific user populations. Significant development effort is spend on fine-tuning the performance of deployed services, especially the call-flow, prompts and speech understanding grammars to reach their target key performance indicators (KPIs). This iterative enhancement and service adaptation process is often performed with little automation by inspecting data log and partially transcribed calls. Thus a clear business opportunity exists for *providing tools and professional services for the enhancement and customization of deployed speech services* both for *service providers* and *end-users* alike. Henceforth we use the term **service doctoring** to refer to the enhancement and customization cycle of speech services.

Recently, a wide-array of technologies have emerged under the umbrella term *speech analytics* that facilitate the automatic or semi-automatic extraction of relevant information from large amount of speech data, e.g., audio mining for key-words and topics, affective analysis, analysis of speaker population characteristics, attitudes and behaviors. The underlying technologies that make up a speech analytics solution include: speech/speaker recognition, emotion recognition, keyword spotting, topic spotting, gender/age identification, call-flow analysis. Speech analytics products for call-centers provide business intelligence related to products, processes and, most importantly, agent performance, e.g., the gopher-it speech analytics platform by Aurix. The system automatically "analyses and evaluates conversations between contact centers and their customers" and provides recommendations for "optimizing business processes and overall customer service quality". For example, the system can identify problematic areas (hot-spots) in the call flow or agents that need additional training. Taking the analogy to the interactive voice response (IVR) arena, one can apply *speech analytic technologies in order to better train the mechanical agent (spoken dialogue system) by enhancing and personalizing the speech interface to better meet user requirements and expectations*. The end goal here (just as is the case for call-centers) is to extract business intelligence and take the appropriate action for boosting the performance of the agent in order to meet the end-user KPIs. By automating the service-doctoring process one can achieve faster service deployment, significant cost savings and improved user satisfaction. We refer to the (semi-) automation of the speech doctoring process by analyzing and modeling large amounts of human-computer interaction speech data, as **spoken dialogue analytics**.

The SpeDial consortium has identified a range of mature technologies as being very relevant for spoken dialogue analytics: speech recognition, call-flow analysis, discourse modeling, emotion recognition, grammar induction, crowd-sourcing and user modeling. SpeDial brings together SMEs that are developing state-of-the-art spoken dialogue systems, tools for tuning speech applications and speech analytics solutions with research institutions at the forefront of progress on spoken language and dialogue processing. Our aim is to *apply these technologies towards the creation of a multilingual service-doctoring platform*. The project result in a commercial service offering for enhancing and customizing interactive spoken dialogue applications that contain two main modules: IVR analytics and speech service enhancement/customization, as well as, a multilingual collection of data associated with the enhancement and customization scenarios.

Jointly with the commercial partners we have identified `low-hanging fruits', i.e., mature technologies that can be transferred to the mobile speech service doctoring platform within the two year time-frame of this project, as well as having an immediate impact on the development effort and quality of resulting speech services. Specifically SpeDial propose to 1) identify hot-spots in the dialogue and propose alternative call-flow structures, 2) select prompt and update statistical grammars for call center applications using transcribed data and 3) identify user population and adapt speech services to their specific needs and

requirements. This is the main contribution of this project; **we expect to save up to 50% of effort/time required to reach target KPIs for deployed services**, significantly improve call completion, and lower the barrier-to-entry for speech services prototyping by introducing our service-doctoring platform.

The main concept behind the SpeDial project is shown next: KPIs and IVR Analytics metrics together with logs/data from the deployed speech application fed into the SDS enhancement/customization module that update/localize the prompts/grammars and dialogue flow of the speech service using machine learning.

An important aspect of this project is *multilinguality*. We have carefully selected a set of technologies that produce state-of-the-art results and are at the same time *language agnostic*, i.e., they do not require deep linguistic knowledge or large amounts of handcrafted linguistic resources (small sets of labeled data might be required for bootstrapping). Using language-agnostic technologies makes it cost-effective to extend the platform to new languages. SpeDial, cover some of the major EU languages, namely, English, French, Portuguese and Spanish, as well as, Greek. Other relatively under-resourced languages such as Turkish, Hebrew and Polish might also be addressed based on SME's business needs. SpeDial **focus on call-center IVR applications** where speech analytics and speech service enhancement using data are most relevant. In collaboration with SME Partners have identified a preliminary list of call-center application domains where the service-doctoring platform will be tested namely: telecom customer service, health and finance.

SpeDial Objectives

SpeDial is built around the *knowledge cascade of technologies, data and services*. Automatic or machine-aided algorithms will be used to analyze the data logs from deployed speech services, and, in turn, these data will be used to tune in a cost-effective manner the speech service using the set of algorithms and tools of the SpeDial platform. The main S&T goal of SpeDial is to *devise machine-aided methods for spoken dialogue system enhancement and customization for call-center applications*.

SpeDial adopts a **user-centric approach** to SDS design. Rather than simply rolling out algorithms from the research lab to the real world (being hopeful about their usefulness), we have tried to map the requirements of a speech services developer and emulate the logical flow being followed. In this process, we have identified two scenarios: **service enhancement** where the developer starts from an existing application and tries to improve KPI performance and user satisfaction, and **service customization** where the developer addresses the special needs of a user population. Thus our second goal is to *create a platform that supports cost-effective service doctoring for those two scenarios: enhancement and customization.* The platform also include interfaces for service and user satisfaction monitoring (IVR analytics component).

Our third goal is to *create and support a sustainable pool of developers* that will be trained to use the platform. Two separate groups of users are targeted: non-commercial users including the research community and speech services developers at end-user companies.

All in all, SpeDial has ambitious but realistic goals both for technological and commercial exploitation of project outputs.

1.2 Achievements and Main Results

Next we outline the achievement and main results in Y1 of the SpeDial project as it pertains to the two main technical goals: speech analytics and service doctoring, as well as the main technological goal, namely the creation of a spoken dialogue analytics and speech service doctoring platform that integrates the research technologies of the partners (i.e., the SpeDial platform).

The main achievement for WP2 and WP3 in Y1 were:

1. Regarding spoken dialogue analytics, two modalities were investigated for the affective analysis of dialogues, namely, text and speech. Regarding text analysis, word-level affective scores were estimated with respect to the three basic dimensions of emotion (valence, arousal, and dominance) exhibiting high correlation with the respective human ratings, e.g., 0.74-0.85 for valence. Several acoustic features were used for the classification of emotional speech with accuracy greater than 80%.
2. A significant part of Y1 WP2 effort was devoted to the detection of problematic parts (hot-spots) of dialogues. Very good results were obtained for all domains and languages ranging from 83% to 90% classification accuracy. Overall, it was observed that the hot-spot detection is a corpus-dependent task and that dialogue act features are really powerful detecting hot-spots. In addition, emphasis was given to two particular aspects of dialogue/speech that are related to hot-spots, namely, user intentions and disfluencies. High accuracy was achieved for the classification of user intentions spanning from 77% to 90%, while 80% (approx.) classification accuracy was obtained used for the problem of disfluency detection.
3. Regarding grammar enhancement, our effort has been divided on two tasks: (i) the development of a web harvesting module for mining and filtering corpora, and (ii) the exploitation of such corpora for building statistical grammars. It was shown that the web-harvested corpora enable the induction of grammars of high quality (e.g., F-measure up to 93%), similar to that of manually collected corpora.
4. The Atelier toolkit was used for the dialogue flow enhancement demonstrating the visualization of dialogue states. The functionality provided by the toolkit can aid a developer in locating problematic spots of the dialogue and correlating them with specific states.
5. Initial experiments regarding user modeling revealed no significant correlation between age/gender and hotspots, while it was observed that elder and females tend to use more words per turn compared to younger and males, respectively.
6. Following the multilingual vision of SpeDial, several experimental datasets were used for the aforementioned tasks, covering five languages. Many of them were created in the framework of the project (e.g., the first Greek affective lexicon), while other existing datasets were annotated according to the objectives of SpeDial. Overall, the features used in the analytics modules were found to be language-agnostic and portable across domains.

The main achievements in WP4 were:

1. Based on the user requirements of the SMEs and their end-customers, the use case analysis, as well as the list of speech analytics technologies available by the research partners of the project we have defined the basic architecture and functionality of the SpeDial platform.
2. Five use cases have been identified based on both the consortium needs for annotated data and, especially the SME business needs. The use-cases have been especially valuable in guiding the platform design.
3. A first version of the platform has been integrated using a hub-and-spoke architecture. The communication is performed via a central hub that acts as a broker between the modules. In addition, the core log processing functionality provided by the SME partners is included in the spoken dialogue analytics part of the platform.

4. A common SPDXml log format was defined by the SpeDial consortium. Each module updates the SPDXml log file with the new information. This way information from the log and analytics information computed from the modules are seamlessly integrated and passed around. The SPDXml log is also used to communicate with the spoken dialogue enhancement module (Atelier).
5. A first version of the dialogue enhancement module was implemented. In the current version of the platform the dialogue enhancement module is mainly used to visualize and identify problematic parts of the dialogue rather than automatically correct them.
6. We have identified the use-cases and real services that we will be working on in Y2, namely the financial services call center that VoiceWeb is currently deploying for a major bank in Greece.
7. The integrated speech analytics modules have been evaluated on various application domains (our five use-cases above) and languages. The end-to-end integration of the platform will happen in Y2.

The objectives of the **dissemination/exploitation** WP5 are to: 1) achieve widespread awareness about SpeDial to all relevant parties (industry, academia, user communities, other EU projects), 2) advertise and promote SpeDial scientific and technological achievements at trade-shows, conferences and other events, 3) exploit the SpeDial platform via user communities, B2B and B2C business models, and 4) manage the SpeDial intellectual property to maximize exploitation opportunities of SpeDial outputs. The main achievement of Y1 are: 1) the creation and population of the project website, 2) the participation of the research and industrial partners in numerous dissemination events advertising the SpeDial project and given demos/presentation, 3) the drafting and first execution steps of the exploitation strategies especially for the SME partners and 4) the management of the intellectual property (especially annotated data) created by the consortium. The main concept behind exploitation strategies of the SMEs which is the central objective of this WP are outlined in the next section.

Overall, as we move to the second year of the project, this report documents our main results and sketches the road ahead. We have built a first version of the platform to provide basic speech analytics and dialogue flow visualization capabilities, an initial hot-spot detection methodology and prototype, and an evaluation protocol to measure the effectiveness of our strategies. In the next year, we are planning to improve on our speech analytics and service doctoring algorithms and platform, as well as commercially exploit the outputs of SpeDial.

1.3 Expected Final Results and Impact

Based on the project plan and the aforementioned progress in Y1, the SpeDial project is expected to achieve all three objectives by project end. The main outputs of the SpeDial project are: 1) The SpeDial platform consisting of two main modules the IVR analytics module and the speech enhancement/customization module. The platform is exploited by the SMEs as a professional service for developers by NuEcho (B2B) and as a speech service offering for call-centers by VoiceWeb (B2C). 2) Multilingual speech services in the telecom, health and finance call-center application domains that will be tuned using the SpeDial platform. 3) Annotated service data that will be made available to the research community (non-commercial license) free of charge for research purposes. 4) Open software for affective modeling and discourse analysis that will be made available by the research partners at project end.

SpeDial addresses an important business opportunity in the area of speech services and speech analytics. Although IVR analytics products and tools for SDS authoring/tuning are available today as separate commercial offerings there is no product that combines the two into a single comprehensive solution. We believe that there is a natural fit between the two technologies, creating a product with strong commercial exploitation possibilities, as discussed next. Speech applications are a natural fit for call-centers helping to improve the user experience and reduce costs. However, all too often deployed services fail to reach their KPIs and achieve the required level of user satisfaction, leading to customer base attrition and reduced revenue. Significant effort and expertise is required to “tune” speech services to reach the required level of

performance (often taking the service offline for tuning). Also tuning speech services is often performed in an ad-hoc way by speech developers going through logs and selectively listening to portions of the recordings. Our proposal for an integrated IVR analytics and service enhancement module for call-center speech applications provides significant advantages: 1) a comprehensive one-stop-shop solution for speech application diagnostics and tuning, 2) on-line tuning of deployed speech services, 3) reduced effort and required level of expertise for service enhancement and customization. Most importantly the proposed solution achieves top performance truly unlocking the potential of spoken dialogue technologies for call-centers. SpeDial's paradigm of rapid deployment of speech services and then incremental enhancement/customization using live data also significantly reduces time-to-market and the barrier-to-entry for speech service providers.

In addition to the aforementioned impacts, SpeDial is also expected to have impact towards the research community by 1) further demonstrating the synergy between speech analytics, machine learning and spoken dialogue development will produce new exciting research and 2) fostering a research community for spoken dialogue analytics using data and open-source software produced by the research partners of SpeDial. Furthermore, providing tools and data for both academic and commercial use will democratize the development of spoken dialogue systems and open it up to a wider audience of developers, as well as, lead to improved technologies for speech services development. It is also expected that SpeDial will have a significant impact towards standards, specifically, in the areas of IVR analytics and the emerging area of spoken dialogue analytics.

For the second year VoiceWeb and Nu Echo will start commercial exploratory activities based on SpeDial's commercial output toward companies in the main domains that the SMEs are addressing –e.g. in the telecom, banking and healthcare sector. The SpeDial output will be integrated in VoiceWeb's and Nu Echo's commercial offerings, while VoiceWeb's and Nu Echo's network of developing partners (mainly SMEs involved in the development of speech applications or resellers of IVR solutions) in many of the countries will be directly involved. The targeted market segment for SpeDial commercial output will be SMEs with live spoken dialogue systems lacking resources to evaluate and evolve them by themselves and current customers with already deployed SDS provided by VoiceWeb or Nu Echo.

Special focus will be given also on social media marketing activities, social media targeting and social media optimization that can reach widely potential in SpeDial interested customers and get their feedback in a more direct and fast way. Social media marketing, can lead to more customers, more traffic, and more conversions, as it is easier to access new customers, and makes a new product more familiar and recognizable for existing customers. Furthermore, social media marketing can yield new business partnerships that will help further disseminate and exploit SpeDial.