



White Paper

The essential role of a speech platform in deploying
effective and reliable speech enabled applications

Authors: Claudia Romellini, Daniele **Sereno** (Loquendo)

Date: September 09, 2004.

It's not so infrequent to simplify the analysis of the performance achievable designing a speech service, making reference to the progress in speech recognition accuracy or to the optimal design of an efficient voice user interface. Speech recognition accuracy indeed has increased dramatically over the last years, but it is still not a perfect technology and there are several many other aspects that can undermine the success of an excellent voice application design even being based on the state of the art speech recognition technology.

Loquendo has gathered a great experience in facing the design and deployment of complete multi-site distributed solutions (from basic technologies to fault-tolerant platforms) for leading customers in their mission critical applications, serving millions of users with complex and challenging voice applications, sometimes with advertising driven, heavy traffic peaks. This rare opportunity has lead to the acknowledgment that only the optimal combination of accuracy and proper architectural implementation of the various elements of a speech platform, allows to exploit the potentiality of the advances in speech recognition, to achieve globally speech service acceptance.

The **speech recognition accuracy** still remain of paramount importance, but this does not simply depends on the technology algorithm behind. In the continuous effort to improve recognition accuracy, we have conceived a recognition algorithm which exploits both Hidden Markov Models and Neural Networks, taking the combination of benefits from both techniques. But this alone is not enough. Low sensitivity to the background noise can be achieved with optimal audio input management at the platform front-end level. The signal distortion due to transcoding is mostly achieved in the network (especially mobile), but it can be increased by internal signal management by the platform if not optimally controlled. Proper detection and cancellation of signal echo is a typical issue of the audio input processing and it is essential in providing best performance. Moreover echoed signals can be either electrical and acoustical, they can be very different in delay and level according to the network characteristics and to the handset characteristics. Non-optimal consideration of these aspects can provide poor performance.

Another important element affecting the appealingness of a voice application is the **reactivity** of the system during a dialog interaction. Experts users typically don't wait until the end of a prompt before issuing a voice command. Fast reactivity of the system combined with a low rejection rate are the right drivers for a good application. But reactivity depends on the implementation strategy for barge-in functionality, on the number and length of buffering introduced in the platform elements and also on the policy adopted in implementing the client-server architecture to host the technologies servers. Another essential issue in this context is the accuracy and robustness of the voice activity detection algorithm. Fast reactivity to noise in the background, interpreted as speech signal, can be more annoying than an acceptable delay in interrupting the prompt but with a higher discrimination between speech and noise.

Even with a high recognition rate and strong noise immunity, a voice application can be frustrated by a poor quality of the prompted messages. In this respect the availability of natural sounding **Text To Speech (TTS)** is another key issue to achieve best results. While the use of prerecorded messages can provide the studio recording quality, this approach has two major drawbacks: one obvious relevant to the higher cost and the second regarding the possibility to easily modify messages for a better tuning of the application. This is of great importance even for static messages that can however require minor modifications along the time and that with TTS don't require additional recordings of the same speaker used to collect the messages. One additional aspect, is the possibility to mix special sounds to the prompt messages, even driven by the service application.

When moving from proprietary IVR to new generation voice platforms, another essential element affecting quality, is the availability of a high performance **VoiceXML Interpreter**, that allows applying to voice applications the advantages of a web-based development and content delivery approach. Indeed the VoiceXML interpreter is one of the core elements of a voice platform. Its efficiency and reliability are responsible to a great extent of the fast reactivity of a voice service and in turn of the quality. Typical issues that can provide great performance differences are the suitability of the cache mechanisms to speed up service pages interpretation, together with the full compliance with the W3C VoiceXML 2.0 Recommendation.

Another key issue in the task of delivering optimal quality of service, is the availability of management tools and standard interfaces and protocols to allow operation, maintenance, reporting and provisioning functions needed to manage deployed services. In this respects, special **administration tools**, should provide those means to easily collect relevant data from the field operation in order to optimize and fine tune the voice application, according to the actual users preferences.

Moreover, in order to ensure to follow technology improvements, thus increasing the quality of the application, when facing upgrade aspects, the **packaging choices** are also important: SW installation must be fast and easy and software upgrade must involve only the changed components ensuring minimum down time during operation as well as retrieval of all reporting data. For instance, when a new ASR language or TTS voice is required, it should be possible to simply install the new ASR and TTS data, and the voice platform should become aware of this upgrade without need of engines or other platform elements installations.

On top of all these issues, the quality and usability of a speech service still depends dramatically on the proper design of the application and of the voice user interface. In this area the **Voice Application Development and Tuning Environment** plays an essential role. Development of a dynamic VoiceXML application means essentially to create a WEB application, so the programming languages, SW application architecture and supporting development environments may be the same used to build any WEB application. The only difference is that the application must

generate a VoiceXML interface instead of a graphical one. So the best support to build voice applications, accessing data over Internet or stored into customer data bases is to use the state of the art of WEB Application Server Architectures and related development environments.

However, concerning grammar development and tuning as well as final debugging and simulation of the VoiceXML dynamically generated service, a different approach has to be taken: these aspects are strictly related to the Voice application domain. In this context, grammar creation/testing/tuning tools and VoiceXML simulation/debugging tools simplify to a large extent the development task. What is really important, for the tuning and testing phase of a voice service, is that such tools provide exactly the same behavior the service/grammar will show during operation. That means grammar compiling and semantic interpretation must be strictly the same used in the platform and VoiceXML debugger/simulator must rely on the same interpreter used in the platform. As a consequence, a close relationship among technologies, voice platform elements and development tools is essential to ensure performance alignment between production and deployment environments.

Loquendo perspective: global quality is a matter of details.

In Loquendo we base our solutions on careful consideration of all these aspects having in mind that the global quality is a matter of optimal design of every single detail. We take care of optimal design of the different technology components involved (see fig. 1) and in particular of the joint interaction of all elements.

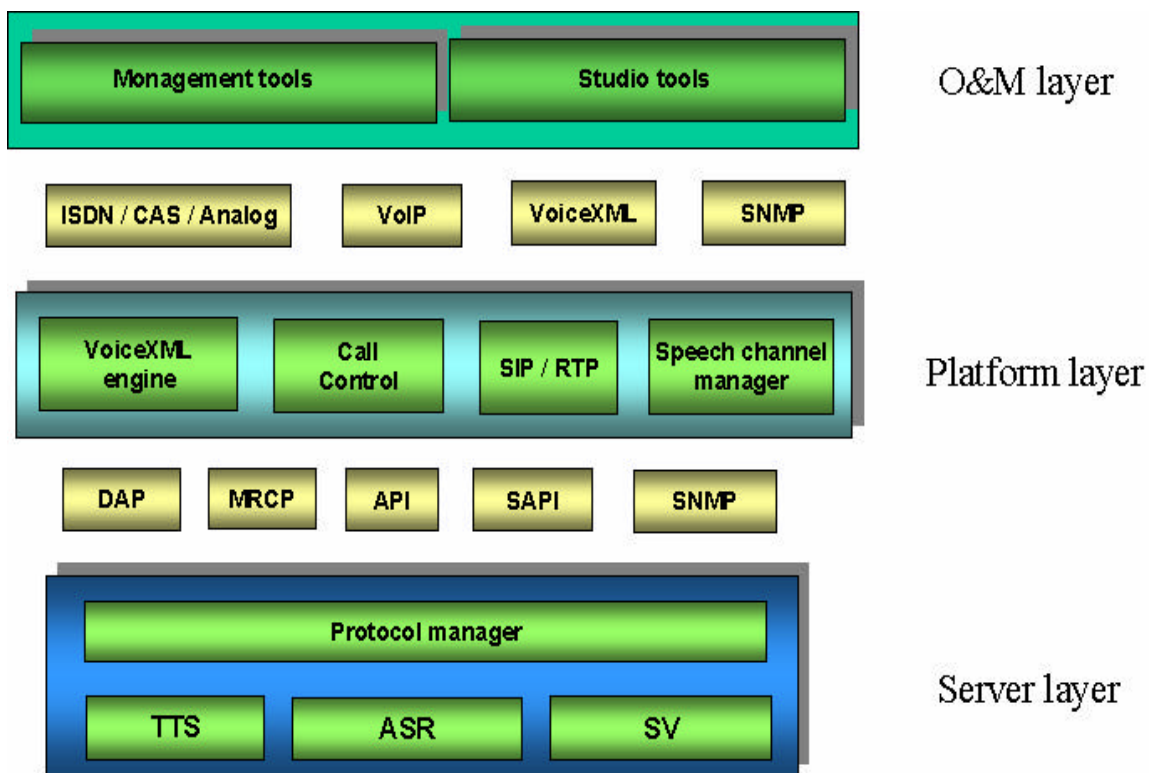


Figura 1– Loquendo solutions open architecture. Loquendo technologies cover all the relevant layers essential in delivering global quality of voice enabled services. Optimal design of technological components is complemented by optimal integration in Loquendo platform solutions.

VoxNauta Lite 6.0, the latest release of Loquendo platforms (see fig. 2), is the result of all these efforts. This flexible, yet complete and easy to use SW voice platform, leveraging on the most sophisticated Loquendo's synthesis and recognition technologies, allows fast deployment of multilanguage VoiceXML 1.0 and 2.0 compliant vocal services.

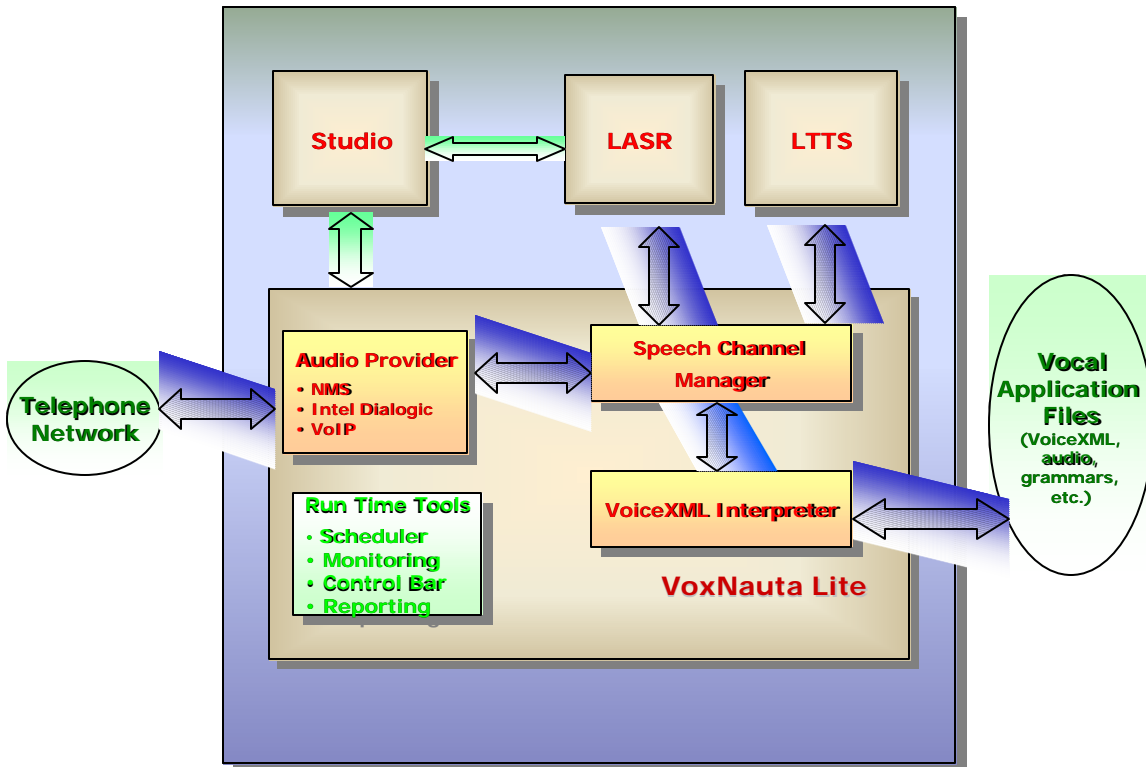


Fig. 2 – VoxNauta Lite functional architecture.

Its renewed packaging gives the customer the possibility to upgrade in a seamless way TTS and ASR engines and to add seamless new voices and languages as soon as they are needed or become available. As it has been mentioned before, this is a key issues to ensure the voice services benefit from the technology enhancements.

The package design, together with a flexible and easy to use configuration tool, allows customer to profile VoxNauta in a wide range of configurations; from a simple IVR to a full dialog platform. VoxNauta, when configured as an IVR, can be used to deploy VoiceXML services based on pre-recorded files and touch tones while configured as a complete full dialog server, it allows deployment of complex voice services based on VoiceXML, TTS, ASR and Speaker Verification. Clearly, between these two extremes, VoxNauta Lite can be configured to work with TTS only or ASR only. The advantage of this approach is to ensure incremental growth of deployed

solutions from classical IVR voice services to full dialog sophisticated applications without losing the incremental effort.

The platform includes a high performance VoiceXML Interpreter, compatible both with VoiceXML 1.0 and 2.0 Recommendations; its high efficiency and reliability has been reached leveraging on the experience gained in large scale deployed systems serving hundred of thousand of calls per day.

Being aware of the impact of management tools on the global quality of the service, VoxNauta Lite 6.0 offers graphical and easy to use interfaces and provide features to monitor and administrate the platform and to collect in a reporting system all information useful for service tuning, such as information about the visited pages and dialog tree .

In addition to the careful attention paid to the combination of the various technologies inside our platform, we have also considered the task to minimize the effort of integrating a platform in his switching environment. In order to facilitate this task, VoxNauta Lite 6.0 offers different profiles to interface telephone network or Voice over IP network, allowing easily integration in a number of different contexts. In particular it supports ISDN NMS boards as well as analog Intel Dialogic and NMS boards. Moreover it offers support to SIP/RTP Protocol for Voice-over-IP network interoperability.

Finally, in the constant effort to easy the deployment task to our customers, any phase of the service development is assisted by the new **Loquendo Studio 6.0**, offering support to grammars development, testing and VoiceXML debugging. The suite of tools is a complete web-based tuning environment complementing VoxNauta and configurable either in single-user mode or as a “software factory” environment, with several development teams working on different software projects.

Conclusions

In spite of the continuous, yet significant, advances in speech recognition, general adoption of automated voice services still depends from the specific service context, on the ability to design the optimal voice user interface in consideration of the service expectations and, to a large extent, on the careful optimization of the various elements affecting the global quality of service.

We have briefly presented a number of issues contributing to the effectiveness of a speech enabled application. All of these interact in a non linear fashion, requiring either the optimization of performance of each component, but also the optimal integration and interworking at the speech platform layer, where technological functionalities are actually exploited.