

Future Landscape for the Speech Recognition Industry

In a few years' time the speech-recognition industry's landscape will be significantly different from the current landscape:

- A consolidation will occur within the recognition-engine companies.
- Speech-application providers will specialise.

The use of open standards is therefore vital.

Engines, platforms and applications

The speech-recognition industry can be considered to be a modular industry: engines, platforms and applications.

On one side, we have the engine providers. These companies provide the engine (or horse-power) to perform the recognition operation. In simple terms, we can consider the engine to be the mechanism that takes a sound wave and a grammar (or a possible word-having-been-uttered) and delivers an ordered list of words that are its 'guesses' at what was said. Companies such as Nuance or SpeechWorks are engine providers.

On another side, we have the application providers (or system integrators). They build on the engine to provide end-user applications. Such applications can be tailored to a company's purpose. Or they can be generic applications, such as the directory dialler. The application developers build the relevant dialogs and perform any required useability testing. Applications also take care of integration with back-end systems (such as directories of names and numbers of employees). This is why they are also considered to be Systems Integrators. In Australia, companies such as VeCommerce, Inflection and Holly are application providers.

There is usually a platform (Interactive Voice Response or voice gateway) that manages the interactions between telephony, engine and application. Such architecture allows engine and application to be independent. Companies such as VoiceGenie, Telera or Nortel-Periphonics are platform providers.

Current strong links between engines and applications

Currently, engine and application providers have relatively strong links. Each application provider tends to use one (and only one) engine.

Such links are more commercial than technical. Application providers act as channels-to-market for engine companies. Normally, enterprises will deal with an application provider and will not worry about the engine in use. However, in order to achieve cost savings through economies of scale and simplified maintenance, big corporations might want to standardise the engines they use across their corporations.

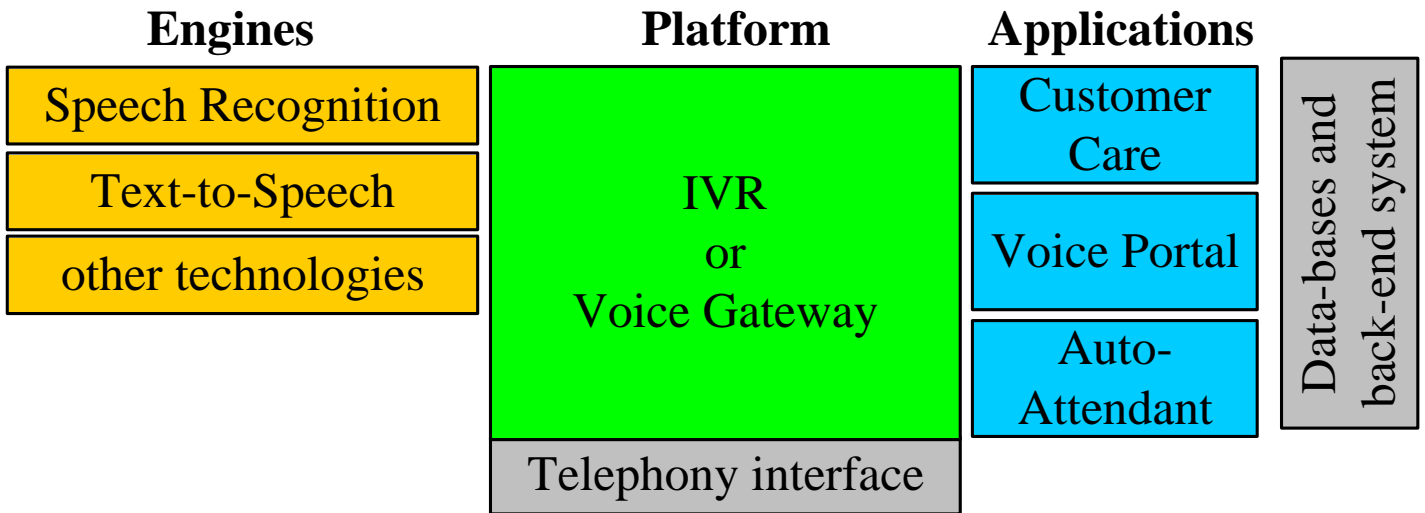


Figure 1. Interactions between engine, application and telephony are handled via a platform

Future unbundling of engines and applications

In addition to the technical architecture discussed above, which allows the unbundling of engine and application, the facts described below support the scenario for an 'unbundled' future.

There will be a consolidation of engine providers. Such consolidation is comparable to the consolidation that occurred in the past in personal computer Operating Systems. As the engine functionality becomes a commodity, it will become harder for engine providers to differentiate their offerings.

Only a handful of engine providers will survive the current phase. This makes it important for enterprises to be able to change easily from one engine provider to another.

A brief look at the evolution of the average share price for some engine companies shows the likelihood of takeovers (due to low share price). One would assume that big corporations would buy niche engine players. However, this is not proving to be the case. For instance, Philips Speech Processing was recently sold to ScanSoft.

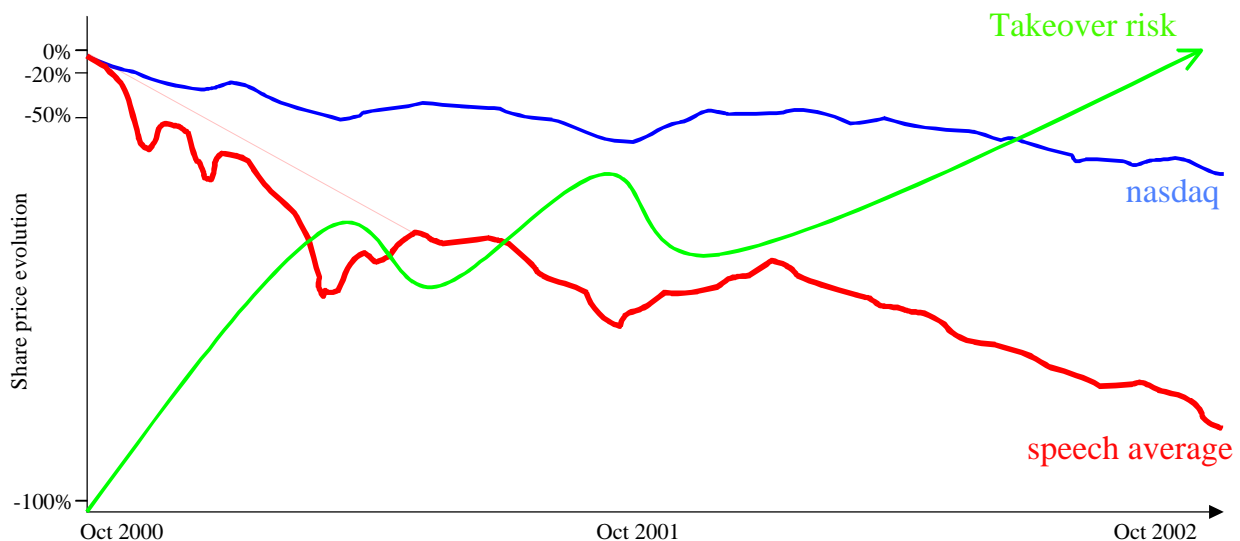


Figure 2. As the share price decreases, the take over risk increases

Application providers will become specialised in application types, such as voice portals or customer care.

Application companies, in order to be successful with such specialised offerings, will need to be able to integrate effectively with any of the engines on offer, even though they have preferential agreements with some of them.

In this future landscape it will become easier for enterprises to change engine providers, and for application providers to integrate effectively with different engines.

Open standards will be vital

For end-user enterprises to reap the rewards of speech-recognition investments, the use of open standards is (and will continue to be) vital.

The use of open standards for communication between the different components (engine, applications, platform, telephony) will facilitate:

- the replacement of one engine by another, which might be necessary in the event of a engine company's disappearance;
- the use of different application providers with the same engine, which will facilitate (and make cost effective) the creation of applications to use on top of a common infrastructure.

The open standards that enterprises should consider, or monitor the evolution of, include:

- Voice eXtensible Markup Language [VoiceXML];
- Call Control XML [ccXML];
- Speech Recognition Grammar Specification [GrXML];
- other World Wide Web Consortium emerging standards, such as Stochastic Language Models N-Gram Specification, Natural Language Semantics Markup Language, Semantic Interpretation for Speech Recognition;
- other Internet Engineering Task Force emerging standards, such as Media Resource Control Protocol [MRCP] and its future derivatives developed by the IETF Speech Services Control group [speechsc].

Survival in the future landscape

Open standards are one of the keys to survival for enterprises and providers in the future landscape of speech recognition, a landscape in which there will be fewer recognition-engine companies, and speech-application providers will be more specialised.

Your comments

This article means to be a provocative trigger for thought. Please send your comments or ideas to me at: jordi_robert@internetaddress.com . I will summarise and acknowledge them in a future note.

Editorial Note

This article is a revision of an original unpublished draft dated November 2002.

About the author

Jordi Robert-Ribes is currently Manager R&D at a top Australian telecommunications carrier. He also works as independent consultant for technology investment venture capitalists. He holds a Post-Graduate Certificate in Financial Management and a PhD in Signal Processing and Automatic Speech Recognition.

Disclaimer

The opinions expressed in this article are the author's personal opinions and do not necessarily reflect the opinions of his employer or previous employers.