



P. Baggia



R. Pieraccini

'From One Speech Luminary to Another'

Speechcycle's Chief Technology Officer **Roberto Pieraccini** (*Speech Technology's Speech Luminary 2008*) and Loquendo's Director of International Standards **Paolo Baggia** (*Speech Technology's Speech Luminary 2009*) discuss the current state of speech - from next year's SpeechTEK Europe to how to change people's perceptions of speech technology.

Roberto Pieraccini: Congratulations on being named Speech Luminary 2009, Paolo! Were you surprised to win the award?

Paolo Baggia: Well yes, I was. It was quite an honour. I saw the award as an acknowledgement of my involvement with the W3C over the last few years, and my longstanding presence in research and in the deployment of speech technologies along with my various academic commitments. And you, Roberto, I hear you were given an academic award just recently.

Roberto: Yes, I was very happy to be elected ISCA Fellow at the INTERSPEECH 2009 conference in Brighton, UK. It is a great privilege.

Of course, INTERSPEECH has a far more research-based slant than SpeechTEK. I actively try to create links between the academic conferences and the more business-oriented ones, and this year I was very pleased to see that there were more people from the academic world in NYC for SpeechTEK 2009, and more people from industry at INTERSPEECH. I hope that this exchange of ideas between the two communities will continue to expand and spread, because they have much to teach each other, and everyone benefits.

Paolo: I strongly support your attempts to bring the two worlds together, and in a similar way I'm working at promoting speech standards within the universities to encourage them to leverage existing and highly effective means of integration, such as VoiceXML and other related standards developed by the W3C and IETF.

Roberto: Given that Loquendo is based in Europe, what are your feelings on the upcoming SpeechTEK Europe 2010? Do you think it will be a success?

Paolo: I hope so, it deserves to be. Loquendo is involved with the committee for setting up the conference, and I think it's a great idea to export a large speech conference over to Europe, where existing events are generally fairly small-scale and tend to be more focused on the technological issues and R&D than on commercial needs. I think a large event will put speech in the spotlight, and so hopefully convince the rather reluctant European markets of its value. How many of the key European players in speech will make an appearance at SpeechTEK Europe remains to be seen – I hope we see most of them there because nothing can beat meeting up face to face, it really is invaluable.

Roberto: I think in the U.S. we tend to think that Europe has rather fallen behind when it comes to the adoption of speech technologies. In North America, speech-enabled apps are by now very mainstream and there is a huge variety of them in use. I read recently that the UK is about a year behind the U.S. in terms of the take-up of speech, with other European countries even further behind. And yet wasn't Europe ahead of the U.S. not so very long ago?

Paolo: Yes, you're right. In telecommunications, for instance, 2G and then 3G mobile penetration started off stronger in Europe than in U.S. and, in the nineties, the diffusion of speech applications was relatively rapid. But then adoption slowed down, and now we have been left a little behind.

Roberto: Why do you think that is?

Paolo: I think there are many socio-economic reasons. For a start, there hasn't been a sufficient level of investment in improving the quality of speech-enabled services, and therefore those that exist are not always as advanced as they might be. Business is often not motivated to invest in speech, chiefly, perhaps, because of long-held European perceptions about the limitations of the technology.

Because Europe was an early adopter of speech, it could be argued that both consumers and

businesses over here have outdated ideas about the quality of speech recognition and speech synthesis, believing it has still not matured sufficiently to justify the expenditure. The message that speech has come a long way, and that it's saving enterprises millions every year, is just not getting through to some European sectors.

Roberto: Yes, exactly, we're not getting that message out. Speech-enabled contact centers, for example, can actually provide a superior service compared with services which are exclusively human-operated. In the U.S. there has been a lot of work, and SpeechCycle is closely involved with this, to demonstrate to customers that automated services *complement* human operators. And of course, operators have considerable training costs associated with them, and that makes speech an attractive alternative. SpeechCycle's aim is to provide an equal or better caller experience compared to human operators. This is actually possible today when sophisticated platforms that embed the most advanced speech science, VUIs, and Web 2.0 concepts are used, like for instance SpeechCycle's RPA (Rich Phone Apps) and Loquendo's VoxNauta (its VoiceXML/CCXML Platform). There has been a great deal of negative press about spoken dialogue systems. The reality, however, is rather different from people's perceptions, in that a modern IVR very often enables you to solve your problem over the phone in just a few minutes; in a traditional contact center you would be waiting on the phone for at least 20 minutes just to get through to a human operator, who, moreover, may not have received the necessary training to be able to solve your problem in the best possible way.

Paolo: What do you think of initiatives like GetHuman.com? Are they helping, or only creating obstacles?

Roberto: I think GetHuman did a good job in spreading the word about which applications were really performing badly. Applications that lock you in and don't give you a way out when the voice interface fails have been poorly designed and are not giving customers enough choices. I think initiatives such as GetHuman came at just the right time and have gone some way towards shifting the focus back onto the caller's needs. On the other hand, GetHuman also contributed to reinforce the negative view that the general public has always had towards speech applications. Speech applications suffer from a lot of negative press, but then it's always easier to criticise than to try and understand. And few people seem to be successfully presenting the other side of the story. We are not getting the message across about how the voice interface can really help people to interact with technology, nor are we explaining the reasons why the technology can sometimes fail.

Paolo: I think the same is true in Europe. There are few attempts to talk about speech technology in an educational context, whereas on the news or in the media in general I often see people discussing and getting excited about other forms of technology. Speech, however, is generally left out of the discussion.

Roberto: We were talking about slow adoption a moment ago, but I think it's interesting to remember that telephone switching was once all done by hand, until automated switching was invented. And the catalyst for that was lack of manpower – AT&T calculated that there simply weren't going to be enough people to do the job, not in the whole U.S.! Then take the ATM machine, for example. When they first came out people didn't really *trust* them, but now more or less everybody *prefers* to get cash from a machine than over the counter in the bank. People understand the advantages of ATMs, but also they accept their limitations. This is not the case with speech. We need to help people to overcome their mistrust. But people *can* be rather intolerant of its limitations - or perhaps we have raised the bar too high.

Paolo: Perhaps it is the continual comparison with human operators, the eternal search for natural, lifelike speech, which creates problems for speech technologies? We mustn't create expectations which can not be fulfilled.

Roberto: Yes, I think you're right. No one's suggesting we abandon the goal of human-like speech, but we should make more use of speech for performing simple, repetitive tasks to a really high standard, so avoiding unfavourable comparisons with human operators. Do you think that the advent of small, handheld devices with a screen will help speech applications to improve their image?

Paolo: Yes, speech can greatly complement other modalities. However, I don't see much common ground or shared features across different speech apps. In the GUI approach, for example, the

interface was simplified by means of double-clicking, icons, shortcut keys, etc. – and these features work on a huge range of different devices, which all share the same paradigm. New devices like the iPhone, on the other hand, are further extending the GUI by introducing touch to enlarge items or move things around the screen. I don't see an emerging common paradigm for speech applications, however.

Roberto: Both ETSI and ITU tried to standardize voice commands, but they were not very successful.

Paolo: We must work harder to update perceptions. When people think of speech technology, they generally have in mind bad contact center encounters from the past. Of course, when people call a helpline it's generally because of a problem of some kind – a broken modem, being overcharged, etc. So they're already irritable even *before* they speak to an operator - virtual or otherwise - and have very low tolerance for any limitations the technology may have.

The truth is, their experiences were already negative even before they picked up the phone and dialled the number. Unfortunately, such impressions continue to reflect badly on speech technology, and they're hard to shift.

Roberto: Yes, and so it also follows that the more speech is found on smartphones and sat-navs, for example, and in the entertainment sphere in general, the more we will manage to shift some of those negative perceptions. And speech works very well in that context because it really complements other modalities like touch, small screens, etc.

I like to think of speech as a situational ingredient – an excellent means of interaction when used in the right place and at the right time, and in the right situation. Interacting by voice might not work so well when you're in a very noisy restaurant, for example, but it's ideal for clearing out your inbox when you're at the wheel. We've simply got to get the message out that speech is *the* most natural way to interact with technology.