Contactcentres.be

Belgische Vereniging van de Call/Contact Centers • Association Belge des Call/Contact Centers
Belgian Association of Call/Contact Centres

Speech Technology in Belgian Contact Centres:

From adolescence to maturity

Version 2.0, June 14 , 2007

Written and published by the ContactCentres.be Speech Technology Workgroup, June 2007

Executive summary

The introduction of speech technology in contact centres is driven by three business factors: customer satisfaction improvement, cost reduction, and revenue increase. High return on investment (ROI) and short payback periods are achievable if the target applications are chosen in line with company strategy, if the right expertise is hired, and if state-of-the-art tools are put to use. This is illustrated with various case studies.

Telephony is and will remain the most commonly used interface between companies and their customers in the years to come. In 2005 one out of three IVR ports shipped were speech-enabled, but this will quickly increase to one out of two by 2009.

VoiceXML has clearly established itself as the standard language for speech-enabled IVR applications, given the large numbers of supporting platforms and tools. Proprietary extensions to the standard are instrumental in integrating speech platforms in the existing call centre environment, whereas standard Internet technology (HTTP, Web Services) enables a swift integration with existing back-office systems for Enterprise Resource Planning or Customer Relationship Management.

Implementations of speech projects share many similarities with classic Web projects, but are peculiar in that user-related issues play a relatively more important role. The language-dependent Voice User Interface and the probabilistic nature of speech technology present specific challenges to speech application designers and developers, which need to be addressed with knowledge and expertise.

Table of contents

Introduction

In December 2005 the ContactCentres.be Speech Technology Workgroup (STW) published the first version of *Speech Technology in Contact Centres: Beyond the Hype*. The goal of this document was – and remains – to help contact centre CXOs and other professionals to (re)discover speech technology in as down-to-earth a way as possible, by focusing on current implementations and opportunities rather than by speculating on future trends. We are now eighteen months later, and it is time for an update. So what has happened since?

Internationally we saw more market consolidation as major players like Nuance, Genesys, Cisco, and Microsoft acquired smaller tool or platform providers and hosters (respectively Bevocal, VoiceGenie, Audium and Tellme Networks) to complete their product and service offering. Worldwide over 7 million calls are handled every day by speech-based self-service, routing and portal applications[1]. In the Anglo-Saxon world, speech technology has been successfully applied in call centres for many years, and in continental Europe, market uptake in France, Germany and the Netherlands has been remarkable as well.

In Belgium, an increasing number of early adopters have implemented speech technology in their customer service departments: this was amply demonstrated at the well-attended ContactCentres.be seminar *Speech Technology: Customer Experiences in Belgium*, held in November 2006 in Diegem. It may be true that Belgian customer service departments have been slower to embrace speech technology than their British, French or German counterparts, but their needs are exactly the same: provide a better 24/7 customer service, at an acceptable cost level. By the way, being a later mover in this respect does not have to be disadvantageous, as it gives companies an opportunity to move forward by learning from other companies' mistakes.

This being said, what does this document have in store for you?

After an overview of the major business drivers for applying speech technology in contact centres, more than ten real-world solution areas are identified where the technology is proving its merit. The succinct market review that follows sketches the distribution of IVR ports in the USA and in EMEA, as it evolves from traditional touch-tone to more advanced speech technology. The document then presents the last decade's evolution towards speech technology standards and commoditisation, but also points out areas where proprietary initiatives still make a difference. This is followed by a brief presentation of implementation aspects and operational models of a speech project. A glossary complements the technical sections. Finally, the last section presents a number of Belgian case studies.

The authors of this document have strived to keep the text readable and neutral by excluding detailed vendor information. To accommodate readers looking for more customer references and local vendor contact information, the Speech Technology Workgroup has compiled a Supplier Directory for the Belgian market as a complement to this document, with individual one-page supplier contributions. On top of this initiative, end-user organisations are kindly invited to share experiences and learn from their peers by joining the Speech Technology Workgroup and/or its affiliate User Group, which is currently in the making. Feel free to contact any STW member for further information and feedback. The list of active participants is added at the end of this document.

[1] Source : Datamonitor : An introductory Guide to Speech Recognition Solutions. Understanding the technology, the vendors and the market. Published in August 2006.

Business drivers for speech technology in contact centres

Revenue increase

Repetitive services like information kiosks are simply too expensive to be handled by contact centre agents. To date such services run on classical touchtone-based IVR systems. Speech recognition, however, may enrich the user interface in such a way that information can be retrieved and presented in various ways. As a direct result more applications will become available, hence driving revenue increase.

Customer retention and satisfaction improvement

Enterprises strive to build customer loyalty through different channels, but in the end the customers decide which communication channel they prefer. Speech technology can be one of these channels, and may offer a competitive advantage. The main condition though is that this be a pleasant experience and not a nerve-cracking one.

As users overall get more exposed to speech applications, they will overcome the current reluctance to converse with machines. This will fortify the two previous drivers.

Cost reduction

Like most industries contact centres are under pressure to reduce costs, while they have to increase the level of service. Approximately 60-70% of the costs in contact centres are labour costs. For low value calls, English and French calls can be handled by offshore contact centres. This is not possible for smaller language groups like Dutch. Potential cost reductions will be accomplished through solutions that offer call avoidance (a.k.a. 'self-service') or call time reduction. Much is expected from speech technology, and the return on investment (ROI) is easy to calculate. Although the technology may not reach the same quality level of a human-to-human conversation, customers are willing to adapt as long as their requests are fulfilled in an effective and timely way.

Indirect cost reductions are a result of the 24x7 availability, and the optimised supply-to-demand ratio for the offered calls.

Agile applications, fast implementation and low total cost of ownership

In a fast moving world the customer expects constantly changing content and more feature-rich applications. As with the Web, new generation speech-enabled applications must be as easy and flexible to implement. If the speech-based solutions are too cumbersome or too expensive to implement and to support during their lifetime, the business case will remain weak.

Leverage existing investments

Many enterprises have invested heavily in contact centre technologies with the expectation of operational savings, but have little to show for the effort because the investment is not aligned with current and near-term requirements. While e-mail and the Web continue to emerge as important channels, more focus must be placed on voice service capability, through which most customer contact activity still takes place.

Real-world solutions

Inbound Call Attendance

Inbound call attendance was the first area in which IVRs have traditionally been deployed. Speech technologies have proven their value to improve the way this service is performed in enterprises and call centres:

- **Dynamic welcome messages:** in addition to the traditional standard welcome message**,** companies may like to broadcast dynamic general or personalized information or alert messages. Using text-to-speech, the content of these messages may easily be updated on-line by enterprise business users, in case of exceptional situations.
- **Enterprise virtual receptionist:** through a company voice directory, the caller may reach a person or a department, just by saying its name, the company's department, or even the type of service or product involved. Call filtering can also be introduced: the called party can accept or reject a call or direct it to a voice mail. In each case, the system manages the relation of the dialogue with the caller.
- **Call attendance and routing in call centres**: manages the first level of contact with the calling customer. Enables to qualify the purpose of the call, which will then be handled either by an automated speech-enabled module, or by an agent. Speech recognition allows a much larger amount of potential call reasons to be managed, with a higher comfort for the caller.
- **Wait time management:** in case the inbound call could not be immediately transferred to the desired person, the position in the ACD queue and average waiting time before being served may be announced to the caller. This allows callers to manage their call in a more flexible way and therefore gives a dynamic image of the company. During waiting time, dynamic flash information messages may again be played.
- **Call back requests**: when the requested service or person is not available, a request for call back may be registered together with the number to dial, the expected time range, and an optional voice message.
- **Voice-Mail:** if the caller wants to leave a voice message to a recipient, it may be recorded by a speech application. This information may be sent by e-mail to the desired recipient, together with an optional SMS notification. In addition, the speech application may store the voice messages in a database and provide a web interface to manage its content.
- **Caller identification:** allows to identify the customer or prospect either automatically from his calling number or from information that he/she provides (customer or contract ID, dial number, zip code, address, names, etc…), which is then searched into company CRM databases or public directories. The preliminary identification of the caller allows the dialogue and menu options to be further personalized.
- **Call transfer with IVR attached context through CTI:** when the call is transferred to the desired person or department, dynamic contextual information is

passed to the recipient desktop, in order to allow a popup display of the customer profile and a summary of the information collected by the IVR.

- **Conference calls:** thanks to the CCXML standard, a standard to manage call control in VoiceXML environment, not only can a speech application be used to attend and transfer a call, it may also organize a conference call between several persons such as an inbound caller and several contact persons in the company (ex: a call centre agent + a back-office expert).
- **Video Conference:** the latest evolutions of the VoiceXML platforms now allow video information to be streamed in addition to audio information. As a result, inbound callers using a third generation mobile phone device may enter a videoconference with the contacted person or call centre agent.

Outbound Dialling

Outbound dialing is an application automating the dialing of internal as well as outgoing calls. In the first case, the general phonebook of the company is use; in the second case, the system identifies the caller and loads the corresponding personal phonebook. When employees call, they are asked to say the name of the colleague or correspondent they want to talk to; the system dials the corresponding number, and connects both parties. Moreover, each employee from a set of different numbers, being their office extension, mobile phone or even private home phone, can access this functionality.

Directory Assistance

As for the module explained above, for very large directories more information than just the name of the person is needed to do a successful search. Often the city name or street name is asked first, before the name of the person is pronounced.

Location Finder

Large companies and administrations often operate a network of points of sales and services like shops, stores, dealers, offices, agencies, brokers, hotels, ATM terminals, garages, petrol stations, service centres, night duties, pharmacies, hospitals, etc. Similarly, tourist offices publish information about tourist attractions, cultural and sports event venues.

As a result, call centres are often asked to provide customers with routine information about those locations, contact information, opening schedules, and types of products or services available. Speech technologies now enable the set-up and rollout of location finders that operate 24/24 in automatic self-service mode, which improves customer satisfaction and reduces call centre operating costs.

A speech-enabled location finder allows callers to get information about a Point Of Interest (POI) they want to visit or contact. A POI can be retrieved through multiple search criteria such as location (country, region, city, or a maximum distance from a specific place), the available products or services, or the opening schedule.

A speech-enabled location finder may include a geographical database covering the target areas with a lexicon of the right phonetic pronunciations of the city and street proper names, as well as the hierarchical organization of locations (country - regions - cities – districts).

Marketing Portals

Enterprises and administrations usually have a catalogue of products and/or services. Communication about products and services is a fundamental part of the organization's marketing plan. Up-to-date information about deliverable products and services can be found on a web site or obtained from a call centre. However, handling those calls for information is a routine job that does not generate any immediate value for the enterprise. This pleads for the automation of this service.

A voice marketing portal accesses the company product database that describes product and services in terms of reference, name, category, features, detailed description, target public, price, availability and delivery. Thanks to the integration of speech recognition, callers can express their search criteria at any time throughout the dialogue process, by telling the system about the name and/or features of the requested products. Product proposals and descriptions are provided dynamically with text-to-speech technology, so that the voice portal always publishes the latest up-to-date information about products in the company database.

First-time callers who don't know how to browse the catalogue by voice can be guided by automated prompts that present the product database contents.

Order taking

Retail companies have implemented plenty of self-service e-business applications over the last years, which enable customers to order products and services on-line over the web.

More particularly, home shopping retail companies, in many sectors like textiles and clothing, cosmetics, food and beverages, flowers, office supplies, sports, culture, leisure, home appliances, electronics, banking and insurance, travel agents, etc., are used to interacting with their customers through call centres.

The VoiceXML standard now makes it possible to implement self-service voice applications for product and service ordering that leverage the enterprise web-based e-business infrastructure. Most recent speech technologies (TTS and ASR) may so be used in seamless integration with the back-office Enterprise Resource Planning (ERP) and Customer Relationship Management (CRM) applications. Typical services provided cover customer identification, selection of order items, options and quantities, and delivery specification and payment.

Order Tracking

Directly related to on-line ordering, the order tracking applications may even more benefit from speech technologies. Although an increasing part of customers are using graphical web interfaces to order products, they still mostly use the telephone indeed to claim about the status of a product or service order.

After due customer identification, the speech application may query the enterprise CRM application or database and provide the caller with real-time information about the current order status (content, status, expected delivery date,)

Agenda management

This module is often used to automate appointments, or to automatically fill in the agenda of a doctor or sales representative, for example.

There are several ways of doing this; these are only a few examples.

1) The system can read to the caller the first available date and time for a next appointment. This can be read back by the computer, using Text-to-Speech software.
2) The caller can than say something like: “ I accept”, or “Yes”, or another confirmation, if he accepts this appointment.
3) If he does not accept this proposed appointment, he can ask for “the next available date”, or propose a “date in the second half of August”.

Depending on how the agenda of a particular service is constructed, or depending on how the dialogue is structured, or the way the agenda is filled up, the dialogue with the Agenda Management can be customized, or structured upon request of the involved parties.

Caller identification and speaker verification

Next to traditional ways to identify the caller (he can be asked for his name, address, etc., or the system can identify the calling number), speaker verification is now available.

During a one-time enrollment phase, callers provide some speech samples over the phone. The speaker verification software creates a voiceprint and stores it. This voiceprint models features of the speaker’s voice, but it is not a recording of speech; in fact no audio samples are stored.

During verification, callers pronounce the requested information such as name or ID number. The speaker verification software compares these live speech samples to the stored voiceprint.

Voice authentication tracks the physical characteristics of more than just the vocal cords, so registered callers with a slight cold will not have any problem getting verified. Only extremely bad colds may have an effect on verification. Potential imposters trying to mimic other people’s voices still have a different voice, so their voiceprint will not match. In this way secured access to sensitive information is not compromised.

Studies, and live deployments have shown that when customers are provided the option to use voice authentication, they will use it! In the US, over 150 million secure, automated calls are being processed annually through voice verification applications in the following domains:

- Account management of financial, benefits or healthcare information
- Field service automation
- Sales force automation
- Corporate dialers
- PIN reset
- Facilities access

Customer care

Customer care is a strategic part of the overall customer relationship, where customers get the opportunity to access and exchange their personal information. Typically, access to personalized information happens through web sites or through calls with live operators.

A growing number of contact centres offer their customers a self-service round the clock personalized access to information. After successful identification, callers may access a wide range of information, such as administrative data (contact information, list of subscribed services, applicable rates, etc.), order status and shipment tracking, consumption reporting, current account status, accumulated fidelity points.

Speech-enabled applications may easily extract this information on-line from the back-office CRM enterprise systems. Text-to-speech technology enables the presentation of these dynamic data in their current status.

If the caller is expected to visit the voice portal frequently, the voice application may save the used search criteria in order to be reused in subsequent calls.

At any time, callers may optionally be transferred to a live operator on simple voice request.

Customer registration (pre-paid card registration)

A very heavy burden on contact centres is linked to the registration process of customers. The legal requirements for mobile telco providers are evolving toward the mandatory registration of the owner of pre-paid card prior to activation. This requirement is already effective in several countries outside Europe and knowing that this may represent a significant increase of customer interactions, some operators have chosen to use speech self-service application for the registration of their customers.

The process can handle the transaction completely without manual intervention and proves to be extremely effective and efficient.

Helpdesk

Resetting expired, forgotten or compromised passwords is a routine part of life for millions of information workers. It’s also a tedious, time-consuming process that complicates already busy schedules, cuts into worker productivity and costs companies millions of dollars per year. Password resets are the second most common reason workers call help desks, accounting for about one in four help desk requests. Other examples are frequently asked questions, or problem reporting and tracking.

Speech applications (with speaker verification) that enable users to reset their own password over the phone, cut the cost per reset and reduce time spent by IT help desk staff on password reset tasks by a minimum of 30 percent. Both local and remote users can quickly and easily reset passwords with a simple phone call.

Phone Banking

By providing real-time account access via speech recognition technology, customers can simply ask the system for information such as, “what’s my account balance?” Customers can also perform financial transactions such as, “transfer two hundred euros from checking to savings”. Customers are in control and will not feel limited by traditional telephone menus.

Speech shortcuts and support for natural language phrases allow callers to stay in control of the system. Touch-tone is also available as a fall back option for touch-tone power users.

Speech is used to acquire account details, account history, funds transfer, stop payment, branch hours (holidays), general information, after hours’ messages, branch and ATM locator and transfer to operator.

One of the first speech-enabled applications in the banking world was developed to trade stocks directly from the phone. Identified callers have the choice to listen to actual share prices, or give sell/buy orders on their own portfolio.

Banks offer these new products to their clients to differentiate their e-business self-service strategy.

Travel information & Booking

When traveling, access to up-to-date information can make the difference between catching a plane and sleeping in an airport hallway. About a decade ago, airport information lines were among the first IVR applications to be speech-enabled in the Unites States. Asking for arrival or departure times by simply saying the name of a city or airport is so much easier than having to type in a partly alphanumeric flight number or some other arcane code.

Likewise, knowing the scheduled departure or arrival time of a bus or train *before* getting to the station makes for a much more relaxed travel experience. Offering this extra phone service on top of existing channels like the real-time information boards is perfectly feasible. The information is already electronically available, so why not exploit it, without overloading the call center with routine questions?

Other typical traveler’s information needs include (static) timetables, ticket bookings, lost luggage, car rentals, hotel shuttles, etc. All of these involve names rather than just numeric codes, and are therefore good candidates for speech-driven automation.

Real-time traffic information

The evolution of the economy towards globalisation dramatically increases our need for mobility, leading to an increasing usage of various transportation means. Growing traffic congestion, maintenance works, severe weather conditions, strikes in public transportation often lead to huge loss of times. In such context, travelling people need accurate real-time information to enable them to adapt their travelling behaviour to current conditions and take appropriate decisions to optimise their time usage.

As the telephone is obviously the privilege channel to inform mobile travellers not sitting at their desks, providing real-time traffic information is obviously a strong use case for speech-enabled IVRs.

Many transportation utilities are keen to provide their customers with real-time information about the availability and schedule of transportation mean they envision to use: airlines, trains, metros, bus and tramways, boats, taxis, road and speedways, …

Callers typically mention start and destination cities, or a line references Depending on the available back-end functionality, the traffic line can suggest the caller to take another route.

Outbound notification

Outbound calls or campaigns are time consuming. Outbound campaigns follow a well-defined structure, making them ideal candidates for speech-enabled automation. Sample application areas are market research surveys, appointment setting, direct advertising, marketing campaigns or fundraising campaigns.

Practical implementations of speech-enabled outbound campaigns are:

- Order delivery confirmation
- Appointment confirmation
- Product surveys
- Event Registration

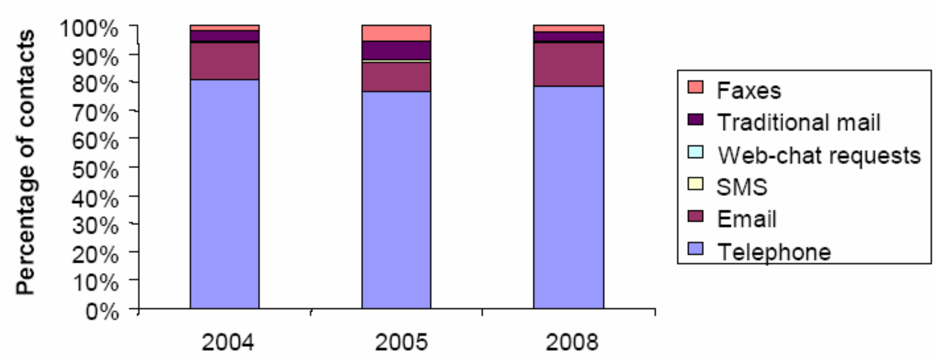
The collected information is stored for further offline handling or can be used in real time when the call is transferred to an agent. The information is used to prioritize and route the call to the best agent.

The timesaving aspect of automated speech can be used in combination with the quality of a real agent. Outbound surveys for example can be introduced by an agent, while the time consuming questionnaire is handled by the speech application.

Global market review

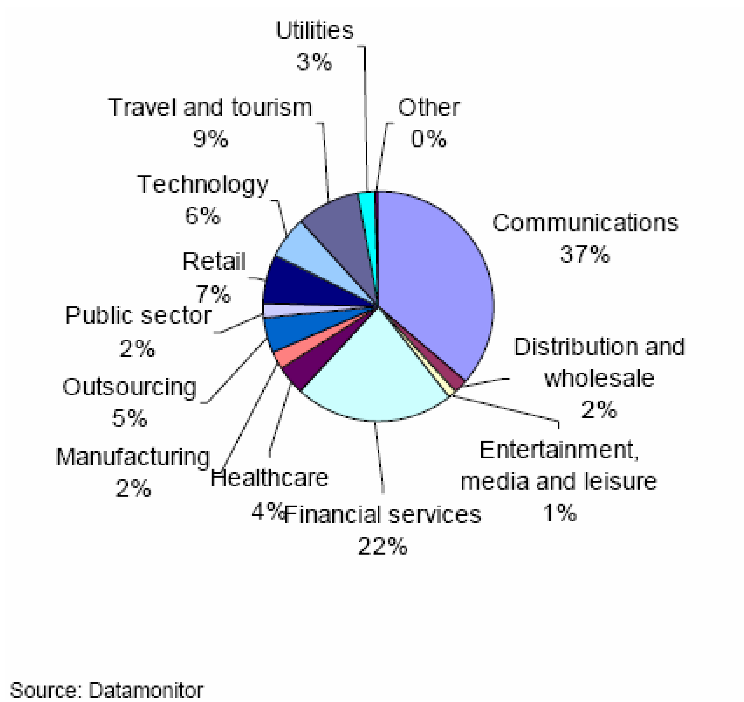
Importance of voice-based self-service

A recent EMEA study from Datamonitor showed that Telephony is and will remain the most commonly used interface between companies and their customers. This clearly emphasizes the need for advanced voice-based self-service solutions.



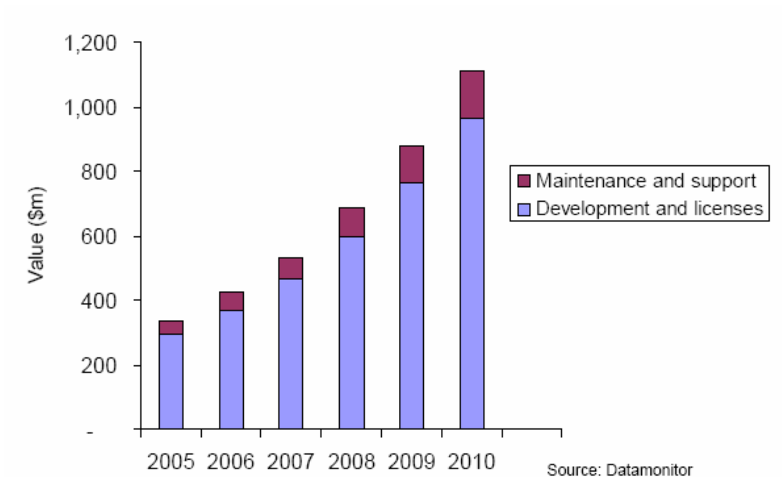
Market segments of Speech-Applications

Early adaptors of speech applications were typically communication companies. Today, we see that speech found his way to other segments market. A recent study of Datamonitor on "Global Market Trends on Speech Applications", published in 2006, included some very interesting information on the global vertical market segmentation of speech applications. The figure below shows the numbers for 2005.



Growth of Speech-Applications

Analysts did significant research about growth of speech application. These numbers provide you with a very good idea about the adaptation of speech in the market. According to recent research performed by Datamonitor, the global spending on speech applications will grow significantly over the coming years at a Compound Average Growth Rate of of 26,8 percent.



Technology trends

According to the analysts, there were 88,000 traditional IVR ports shipped in EMEA in 2004. By 2009, traditional IVR port shipments are expected to dip to 81,000 at a negative compound annual growth rate (CAGR) of -1.6%. VoiceXML shipments will grow at a strong CAGR of 24.9% through the next five years, from 46,000 ports in 2004 to 140,000 in 2009. These numbers clearly confirms the trends toward open VoiceXML based platforms.

IVR port shipment proportions, 2004 & 2009

Ports (thousands)	2004	%	2009	%
Traditional IVR	88	65.7%	81	35.8%
Voice-XML	46	34.3%	140	61.6%
SALT	0	0.0%	6	2.6%
Total	**134**	**100.0%**	**226**	**100.0%**

Technology standards and components

Platforms

Typical legacy IVR and speech recognition platforms are proprietary, closed and tightly integrated systems, operated by scarce experts. They are expensive to maintain, and tie the customer to the vendor through lack of interoperability with other systems and high switching costs. The advantage of having one single contact for support issues does not outweigh the cost of vendor lock-in.

The advent of the W3C-backed VoiceXML standard since 1999 has led to a proliferation of VoiceXML browsers, platforms, and development tools. The VoiceXML Forum, which promotes the language, has more than 150 member organizations. Thousands of VoiceXML developers have developed tens of thousands of speech-enabled phone applications worldwide, the larger part in the English-speaking countries. The VoiceXML standard is now at version 2.1, with version 3.0 being in preparation. There are literally dozens of VoiceXML platforms on the market.

Integration with speech recognisers is nowadays performed via a standard MRCP (Media Resource Control Protocol) interface. Opposite to proprietary connectors, this standardized interface makes it very easy for the voice platforms to support new releases of speech recognition software from different speech vendors.

The Contact Centre industry has seen an evolution in the last ten years. The latest evolution is to provide video as an interaction to the customers in order to personalizing the interaction and improving the quality of the call resolution. Most vendors of speech platforms have already included video capabilities in their product such as video-on-hold and video parking functionalities. This is the capability to play a user-defined file when the customer is put on hold or the call is parked. Typically, a video self service platform provides varied choices and flexibility in the supported codecs (H.261, H.263, and H.264) and file formats (wmv and mpg).

Speech Engines

Text-To-Speech (TTS) or **speech synthesis** systems have substantially gained in quality over the last few years, to the extent that in some application settings they are indistinguishable from a human reader. The technology is particularly useful in a dynamic environment, where the information to be presented to the caller is inherently unpredictable.
Thanks to this improvement in terms of quality, users of TTS wish to have an exclusive voice, which makes it possible to personalize the services. This personalization is possible not only on basis of an exclusive voice but also by including functions like a music background integrated in the TTS.

Systems for speaker-independent **Automatic Speech Recognition** (ASR) have also improved over the last few years, but still require a fair amount of tuning for optimal performance and accuracy. This is especially true in a mobile setting, in cars, or in other noisy environments. The inherent probabilistic nature of ASR systems necessitates well-designed error recovery strategies at the application level. This is where Voice User Interface Design comes into play.

State-of-the-art ASR systems are capable of recognizing utterances from grammars containing tens of thousands of entries.

Systems for **Speaker Verification (SV)** have been deployed to recognize and authenticate callers by their voice. For added security, speaker verification can be combined with other, more classic methods like passwords, pass-phrases or PIN codes.

Tools for Application Development

The simplest tool for developing a VoiceXML application is a **text editor**. More than a decade ago, the first websites consisting of static HTML pages were also developed this way.

To make their lives somewhat easier, developers have written **libraries** or **plug-ins for VoiceXML code generation** in various computer languages and integrated development environments. Although this kind of voice application development still requires a programmer acquainted with the programming language at hand, it's already a step ahead.

The most advanced systems for voice application design, development and management **are fully graphical, and often web-based**. They no longer require programmers to design and develop the front-end of a speech-enabled phone application. This means that more time can be spent on the VUI aspects, which is of utmost importance. Voice application management functionality includes operational statistics, archiving, live monitoring and exploitation.

Integration

One of the big advantages of the VoiceXML standard is that the speech application (and its development) has become totally independent of the underlying hardware. State-of-the-art generic VoiceXML platforms support a wide variety of telephony boards and VoIP connections. Just like HTML pages can be viewed in different Web browsers (MS Internet Explorer or FireFox), VoiceXML applications can in principle run on any VoiceXML platform (although platforms do sometimes have their own extensions or peculiarities, just like with MS IE and Netscape).

A weaker point of these generic VoiceXML platforms, however, is their limited integration with existing call centre software. This is where proprietary extensions to the platform can make a difference. The major call centre software providers extend their VoiceXML platform with a software layer that integrates seamlessly with their existing **CTI technologies,** e.g. for assisted service. When necessary, operators are perfectly able to follow up on a speech-initiated call. This way speech technology and human-assisted service work in harmony to provide a continuous caller experience.

Internally and externally, the speech platform typically communicates in the Hypertext Transfer Protocol (HTTP), the lingua franca of the Internet. The usage of **Internet protocols and technologies** in a speech application setting makes it easy to integrate the new speech-enabled phone channel to existing back-ends that are already based on Internet technology. Existing business logic can be reused, and past investments in technological or human resources are therefore protected.

For example, many Contact Centres already have a **CRM system** in place, which is accessed via a web interface. They can now tap an automated voice channel directly into the existing

CRM back-end, with or without additional human intervention, as desired. Potential issues arising from the voice channel can be trapped and escalated to a human operator.

Operation Monitoring Tools

As speech-enabled IVRs provide more services as traditional IVRs, they have a larger impact on the enterprise business operations. Business users subsequently need accurate monitoring and reporting information to continuously improve customer satisfaction based on a detailed understanding and analysis of the way they use the speech applications

Typical requested information include

- **System monitoring:** a tool for technical users that continuously monitor system hardware and software resources involved in the IVR system, provide a user interface showing all component status, and generate alarms in case of incidents

- **Application monitoring:** a tool for operation managers that tracks the managed calls in real-time by providing real-time information on the handled calls, such as volume, status, breakdown by category, etc … to enable them taking appropriate business operation decisions

- **Historical calls logging:** a tool for quality managers that tracks all call sessions, with a detailed Call Data Record (CDR) for each call including the list of all time-stamped events (such as the followed path in the IVR script, the played messages, the caller entries in DMTF and speech recognition with registered confidence level,)

- **Queries and statistical reports:** a tool for business users that generates configurable statistical reports automatically or on demand. Statistical reports present calls statistics over periods of time, such as call origin, duration, requested services, abandon and error rates, … They enable to track the evolution of the customer's behaviour over time. The IVR statistical reports often need to be consolidated with the overall contact centre reports to provide accurate business intelligence analysis.

These tools are typically included in the Voice Application Management Systems.
The user interfaces need to be web-based when the monitoring tools are hosted by a service operator.

Implementation aspects

Speech project characteristics

Setting up and implementing a speech project is no different from a classic Web project, except for:

- the increased importance of the (Voice) User Interface, which must follow linguistic and psychological principles of conversational interaction
- the increased importance of usability testing
- an incremental design and development approach, with multiple feedback loops

These aspects are particularly important and most of the speech deployments that have failed in the past were due to a poor design or a design that did not take the user perspective into consideration. When thinking about designing an application, it is important to understand the target audience, their needs and expectations. Who are the callers, what do they call about, what and how much do they expect to find out, how often do they call, how do they phrase their questions, and what terminology do they use and understand? There are a number of specific tasks that can help answer these questions, including:

- transcription and language analysis of recorded 'user-to-agent' calls
- agent interviews
- call centre visits
- listening to live calls
- call-type analysis.

During the design phase, linguistic expertise is invaluable, particularly for designing the dialogue – i.e. the recorded prompts that the users are required to respond to. Good dialogue design is not easy; it needs to be simple, straightforward, consistent, unambiguous, helpful and representative of the language the target audience uses and/or is familiar with. Most importantly, dialogue design must help users navigate their way through the system smoothly. Dialogue is normally designed to reflect a persona – a carefully constructed human image that the company wants to portray to its audience.

Typical project phases are:

- Business analysis: define business drivers, needs and goals; select the right application
- Requirements definition: what are we building
- VUI and Call Flow Design: how will the dialogue between caller and system flow?
- Back-end design: how will the speech application integrate with existing back-end systems (ERP, CRM, CTI, etc.)
- Application development: production of VoiceXML or SALT code, speech grammars, pronunciation lexicons, and system prompts (pre-recorded audio files)
- Testing: usability, functional, unit, load, and integration testing
- Deployment: accept live caller traffic on the speech application
- Evaluation: measure speech recognition accuracy, task completion rate, automation rate, customer satisfaction, ROI or other key performance indicators
- Post-deployment tuning: listen to real caller utterances, compare them with what was recognised, and tune the VUI, grammars, pronunciation lexicon or system prompts for better performance
- Operational maintenance: monitor the live application

Operational models

Depending on the physical location of various system components, we distinguish between three models.

In the **insourcing** model, the company offering the speech system keeps all components in-house. These include: back-end data, speech application, speech and telephony platform.

In the **ASP** (application service provider) model, all components are outsourced to an external service provider, including the speech application and even the back-end data. Examples of ASPs in Europe are BT, Telefónica, T-Com or Telesonera.

In the **VSP** (voice service provider) model, the speech and telephony platform is hosted by a third party, but the speech application and certainly the back-end data remain at the company's premises for added security and control. Each live call generates HTTP traffic between the VSP's speech platform and the customer's Speech Application Server and/or back-end databases. Voice Application Hosting Providers have made it possible for anyone with a web server to develop and serve their own speech-enabled phone applications, in more than 15 languages. Most VSPs also offer application hosting, but more and more customers prefer to keep their application insourced, often for security reasons.

Expertise requirements

The creation of a speech application requires specialised knowledge and skills of various kinds:

- Project Manager
- ICT Engineers (telecom, network, security, provisioning, installation)
- VUI designer (VUI design, call flow design, usability testing)
- Speech engineer (grammar and lexicon development, speech application tuning)
- Sound engineer / Voice Talent (system prompt recording)
- Speech application developer (call flow implementation)
- Web architect (back-end design), Web programmer (back-end API development), CTI programmer (CTI integration)
- QA engineer (test plan writing and execution, bug report writing)

Potential pitfalls

Speech recognition and speaker verification are probabilistic technologies by nature, which means that the speech engines sometimes get it wrong. Knowledgeable application designers and developers tackle this challenge by employing error recovery strategies, and by optimising recognition rates during the post-deployment tuning phase. In case of continued speech engine issues, an application can always back off to DTMF recognition, or transfer the call to an operator.

End users need some time to get acquainted with new technology, and speech-driven phone applications are no exception to this rule. Well thought-out interaction design must guarantee callers a pleasant and effective user experience, e.g. by providing guidance and help when needed. Usability testing and phased rollouts help application developers detect imperfections in an early stage.

Given its cultural dimension, speech technology obviously has strong localisation requirements. Apart from pre-built language-specific pronunciation lexicons and tools for the automatic generation of phonetic transcriptions, third parties such as TeleAtlas have invested heavily in making geographical data ready for speech. Pre-built lexicons with person data (transcriptions of tens of thousands of first and last names in Belgium) are also available on the market. This way, new applications don't have to start from scratch, which significantly reduces development efforts as well as time-to-market.

Glossary

Dialogue

- Call flow: The structure of how a system branches based on callers' responses.
- CCXML: Call Control eXtensible Markup Language. Provides telephony call control support for speech applications (whether based on VoiceXML or not).
- MRCP: Media Resource Control Protocol. Standard interaction protocol between server components of a speech platform.
- SALT: Speech Application Language Tags. W3C-approved set of extensions to existing mark-up languages that enable multimodal and telephony access to information, applications and Web services from PCs, telephones, and wireless personal digital assistants (PDAs)
- VUI: Voice User Interface. Set of interaction elements of the speech system, which drives the caller experience. Includes the "sound and feel" of the application. Crucial success factor of any speech application.
- VXML = VoiceXML: Voice eXtensible Markup Language. An XML-based document format for describing an automated dialogue between a caller and a system. VoiceXML is to a speech-enabled phone application what HTML is to a web application. W3C-backed industry standard, widely supported.
- VoiceXML Browser = VoiceXML Interpreter: server software that interprets VoiceXML code, and manages the automated dialogue between caller and system. Central component of a VoiceXML platform.
- VoiceXML Platform: set of server-based software components, typically consisting of a VoiceXML browser, an ASR server, a TTS server, and various file caching servers.

Input (caller to system)

- ASR: Automatic Speech Recognition. In a telephony context, modern ASR engines are speaker-independent, which means that they do not need to be trained by individual callers before usage.
- Barge-in: The ability for a caller to interrupt a system prompt before it has finished.
- DTMF: Dual-Tone Multi-Frequency. Also called touchtone. In a speech application context, DTMF or touchtone-based input contrasts with input based on the human voice.
- Grammar: Set of rules defining what the recognition engine is able to recognize in a specific dialogue state. Popular grammar formats include GSL and SRGS.
- GSL: Grammar Specification Language created by Nuance Communications. Proprietary but popular format for specifying ASR grammars, supported by the Nuance ASR engine.
- NLU: Natural Language Understanding. The ability to understand complex caller input spoken in a natural, free-style manner.
- SRGS: Speech Recognition Grammar Specification: XML-based standard for specifying ASR grammars, supported by various ASR engines. Vendor-neutral, backed by W3C.

Output (system to caller)

- Audio file: Digital sound file that the computer plays to a caller. Contains (part of) a system prompt, or an earcon, or a mixture of both.
- Call script: A list of prompts to be recorded by a voice talent.
- Earcon = audio icon. A short, sometimes musical, sound.
- TTS: Text-To-Speech. Technology whereby a computer pronounces previously unseen words and sentences. Also known as Speech Synthesis.

- SSML: Speech Synthesis Mark-up Language; XML-based and W3C-backed standard for marking up running text with pronunciation clues, e.g. “emphasis”. Supported by various TTS engines.
- System prompt: A sentence played by the system to the caller. Can be recorded by a voice talent, or played automatically by TTS.
- Persona: standardized mental image of a personality or character that callers infer from the application’s voice and language choices. Persona design is a part of VUI design.

Input and Output

- Lexicon: a list of words and their phonetic transcriptions. Used by ASR as well as TTS systems.
- Phonetic transcription: formal representation of the pronunciation of a word

Belgian Case studies

STIB - MIVB (Belgium)

Challenge

STIB, the Brussels public transport operator, was determined to assist travelers by providing real-time bus and subway arrival times. At first, they simply offered web access to timetable and delay data as reported by the global positioning system (GPS) devices installed on each vehicle. While this service was great for people sitting at home or in the office, it didn't help people who were on the go - i.e. those looking to catch a bus or waiting for one that had been delayed. To meet the needs of mobile travelers, STIB wanted to provide anytime/anywhere access via telephone. The cost of providing phone access first appeared prohibitive, however, because the revenue generated by a ticket sale did not cover the cost of handling a local call. Therefore, if STIB were going to offer this service, they needed to restrict the cost of development and delivery.

Solution

A voice user interface was designed to give callers fast, easy access to the information they need. A rapid development environment was used to create the call flow, develop the application logic and communicate via Web Services with the Web-based locator application that determines the arrival times of the buses. The platform's native support for Web Services protocols such as SOAP enabled direct integration with STIB's IT systems. This eliminated duplicate programming work, and enabled a truly cost-efficient solution.

Results

Now, travelers can call STIB (in French or Dutch), say the bus stop name and line number, and hear the exact arrival times for the next two buses in either direction. STIB reached its goal of providing anytime, anywhere access to schedule information within the financial limits established for the service. People on the go in Brussels can easily get around the city and avoid the frustration service delays can cause.

ATOS Wordline SA/NV (new company name for Banksys and Bank Card Company)

Challenge

Atos Worldline has a worldwide reputation for its electronic payment solutions, hi-tech terminals and security technologies that are approved and supported by the banks, as well as for its extensive range of services. The Atos Worldline's call centre receives more than 1,700,000 calls per year, made by a wide range of clients, card owners as well as from retailers and banks. Until 2006, the inbound calls were attended by a legacy traditional IVR system. Banksys recently decided to change this IVR in order to enable much faster management of the evolutions required by its business, while leveraging the capabilities of new speech technologies to deploy new automated self-service transactions.

Solution

Banksys selected a new platform to respond to the main following requests:

- User interface to be used by non-technical staff from the business group.
- Support of text-to-speech and voice recognition technologies allowing for a self-service processing of routine tasks.
- Compliancy to VoiceXML standards.
- Integration to Atos Worldline information systems (SQL access to an Oracle database and transaction with a Tandem system through IBM Websphere MQ Series).
- Statistics and reporting tools allowing the continuous control of service quality
- Voice call Back standard services and management of service business hours

Results

After a progressive migration of existing applications, within the requested 4 months time frame, Atos Worldline users can now take advantage of a platform that allows them to secure the evolutions of voice server applications proposed to their customers. A first self-service application was successfully put into production. This application is closing calls related to card blocking by automatically giving specific information to the caller. The first feedback inquiries to the users of this new application show satisfactory results. This encourages Atos Worldline in the pursuit of the use of the new technologies to offer a growing number of self-service applications to its customers. Atos Worldline is indeed now considering the progressive deployment of voice recognition to enhance the interaction with the callers.

DEXIA (Belgium)

Challenge

Dexia counts hundreds of high net-worth clients who regularly make a simple phone call with a Dexia Direct Private (DDP) trader to get accurate trading information and give buy/sell instructions. Originally, VIP clients had to authenticate themselves with each call by keying in a lengthy contract number and PIN code. After a while, this process was considered an unacceptable nuisance. DDP was therefore looking for a user-friendlier authentication solution, which at the same time would maintain the highest security standards.

Solution

In merely three months, a consortium of primarily Belgian technology partners designed, developed and installed a speaker verification solution for Dexia Direct Private. The solution integrates speech recognition, speaker verification and CTI technology. Here's how it works: after a one-time DDP-initiated enrolment phase, VIP callers are asked each call to simply say their name, and answer a secret question. They are then transferred to the DDP professional, who gets a clear on-screen indication of the caller's authentication status. The authentication process is fast, simple, and efficient, as the VIP client does not need to remember a contract number or PIN code anymore.

Results

After 6 months of operation, hundreds of VIP clients were (and still are) successfully using the speaker verification system. The solution was not only directly adopted by the VIP clients, but also by the DDP traders. Apart from the technical factors, the solution is a success because it appeals to the autonomy and self-sufficiency of the typical VIP client, who is naturally receptive to the concept of a personalised, high-tech, and high-security authentication method.

Contributors

Marc Acoe (Fortis Bank, STW member)
Patrick Cauderlier (Acapela, STW member)
Ludo Cox (Genesys, STW member)
Jurgen De Becker (Genesys, STW Member)
Roel De Geus (Nuance, STW Member)
Frederik Durant (Prompt! Speech Applications, STW member)
Peter Edel (Quentris , STW member)
Laure Rasson (Nespresso, STW member)
Alain Rondenbosch (DB Scape, STW chairman)
Isabelle Simal (Nespresso, STW member)
Wim Van Cleemput (IP Globalnet, CC.be and STW Member until Oct. 31, 2005)
Willy Van De Voorde (Ubicall, STW member)

With assistance from:
Jean-Marc Pajot (Dimension Data, CC.be Technology Steering Committee Chairman)
Vincent Vanden Bossche (General Manager ContactCentres.be)

Contact information :

For any question or suggestion on the content of this document, please contact:

Alain Rondenbosch
Chairman
Speech Technology Workgroup
ContactCentres.be
e-mail : rondenbosch@dbscape.com
Tel : +32 2 354 10 96
Fax : +32 2 354 64 47

An electronic version of this document as well as the Speech Technology Suppliers Directory for the Belgian market are available on-line on www.contactcentres.be