Zeitschrift: Comtec : Informations- und Telekommunikationstechnologie = information and telecommunication technology

Herausgeber: Swisscom

Band: 79 (2001)

Heft: 9

Artikel: Voice service design : do it yourself

Autor: Marti, Urs-Viktor / Van Kommer, Robert / Krone, Oliver

DOI: https://doi.org/10.5169/seals-876567

Exploration Programmes:
Corporate Technology Explores Future Telecommunications

Voice Service Design – Do it Yourself

The aim of open voice services is to give customers the possibility to create their own speech-enabled service, allowing the end user to navigate through the service with his/her voice. At the moment many of these services use either human operators or DTMF; only few are based on speech technology, either because it is expensive to develop such applications and/or no acceptable dialogue can be found. In this article we present our vision and solution to create a speech-enabled service in an easy and fast way. The realisation combines common internet techniques with state-of-the-art speech technology.

The Programme "Open Communication Services Architecture" develops person-to-person and person-to-content communication services based on an open communication service architecture. It concentrates on advanced real-time services for Open Switch communication systems and makes use of state-of-the-art technologies such as Voice over IP, Web, WAP, Intelligent Network technologies, as well as enhanced media processing technologies like voice processing and Intelligent Agents.
With its Exploration Programmes, Corporate Technology is exploring telecommunication technologies and new service possibilities with a long-term view of 2–5 years. Further, the expertise built up in the course of this activity enables active support of business innovation projects.

Using voice is the most natural way for humans to exchange information. Hence, it seems natural to use spoken language as a man-machine interface for various services. Today, first

URS-VIKTOR MARTI, ROBERT VAN KOMMER AND OLIVER KRONE

speech-enabled services are reaching the market. These services are created individually with large efforts in time and money using complex development cycles.
In order to reduce time-to-market we use an approach based on Internet technology, which facilitates the design and deployment of speech services significantly. The aim is to make speech service development as easy as developing a standard webpage by providing an Internet based speech service creation environment ("open voice service"). This means that the service description is split from the telephone platform hardware. More specifically, the service description may be located anywhere on the Internet. This is possible by expressing the service description using an open standard of markup language: VoiceXML [1].

Voice is the most natural way for humans to communicate. Therefore it is desirable to use voice as the preferred man-machine interface to access various services. However, the creation of speech enabled services is a very complex task demanding huge investments of the service provider. Significant cost reductions can be achieved by providing an Internet based speech service creation environment which allows the speech service provider to interactively – by using a web interface – design ("click") and deploy the desired speech service.

A lot of Internet users create their own, very personal homepage without knowing how to program HTML. This is possible due to tools and layout templates for webpage design. The same concept can be applied for the creation of speech-enabled services. Customers can choose from a certain number of simple and often used dialogue patterns to build their own individual services. This can be achieved by combining VoiceXML (a markup language for voice services) with standard web technologies. While VoiceXML gives access to speech technology, web technologies enable the creation, update and launch of speech services by using a common Internet browser.
This concept will reduce the amount of time and money for Swisscom customers, as well as for Swisscom itself, to implement these services. By combining the "open voice service" concept with a portal to sell the necessary service access number (typically a 0800 number), Swisscom will be able to offer a new attractive service to its customers making speech-enabled services available for a wide range of applications and customers (fig.1).

Open Voice Services

As part of the exploration programme "Open Communication Services Architecture" a system has been developed which allows the customer to create his/her own speech-enabled service within a couple of minutes. The system is built on a set of webpages which guide the customer through the creation and administration of a speech-enabled service. It offers the components needed to run a speech-enabled service: a speech recogniser and synthesiser, commonly used dialogue patterns, and basic administrative functionality.
All service components are interactively combined using a standard web browser. The completed service components are then automatically stored on a web server and the corresponding VoiceXML description is generated. Upon service launch the VoiceXML code is uploaded and interpreted on a VoiceXML gateway. From this moment on the service can be used over any phone (fig. 2).

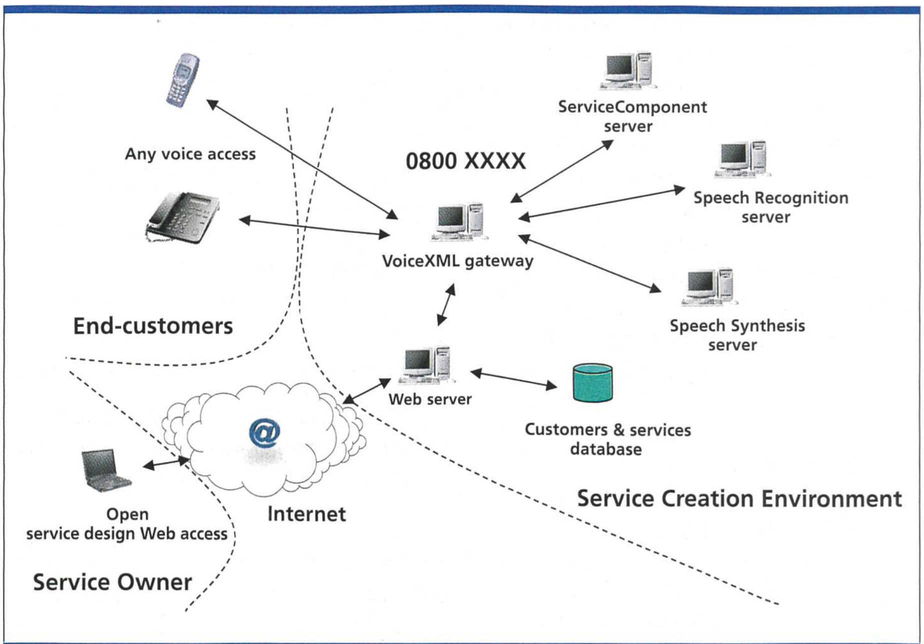

Fig. 1. Open voice services give the service owner the possibility to create, configure and maintain his own speech-enabled application over the web. The end-customer can access the service over any phone and navigate through it with his/her voice.

Speech Recognition

Recognising words and sentences spoken by users is one of the key technologies in this concept [2]. Of course, bad recognition of the spoken word will lead to bad acceptance and usability from the user's point of view [3] while good recognition performance supports the dialogue design [4]. In order to reach good recognition rates, tests have shown that two constraints have to be fulfilled: first, a rather small vocabulary with distinctive words is needed and second, a good dialogue design helps to lead the user to the expected answers. If these constraints can be fulfilled, a certain guarantee for recognition success can be given.

Speech Synthesis versus Recorded Prompts

Two methods are common for producing the spoken language output. The first one is to record texts spoken by a person. These records can be replayed during the dialogue. The second and more complicated way is to generate spoken text from a given text artificially [5]. At the moment, such generated prompts still sound artificial, and thus text spoken by a real person has higher user acceptance. On the other hand, for dynamic information, which may change very quickly, synthetic prompts are needed.

Dialogue Patterns

The third important aspect for speech-enabled services is the dialogue design. A well-designed dialogue can help the user to reach his/her goal efficiently and will lead to a good acceptance of the service [6]. Because the customer rarely is an expert in dialogue design, we have made available a certain number of dialogue patterns, i.e. a message player, a voice menu or a yes-no question. These patterns are optimised in such a way that a good dialogue is possible with only few parameters given by the customer.

The Voice Menu

As an example of a commonly used dialogue pattern we have realised a so-called voice menu which can be configured and maintained over webpages just like any other component. With this dialogue pattern a choice of a certain number of possibilities can be implemented. First, a welcome prompt has to be defined or uploaded. Then the actual choice functionality has to be realised by defining menu key words. For the example of a speech-enabled weather forecast service the key word "info" will provide additional information about this service, and "weather" will lead to a submenu in which the city may be chosen. This submenu is realised in the same way as the main menu. In the operational phase of the service the key words have to be spoken by the user to get the appropriate information. Each menu item (in our example "info" and "weather") can be linked with other speech components

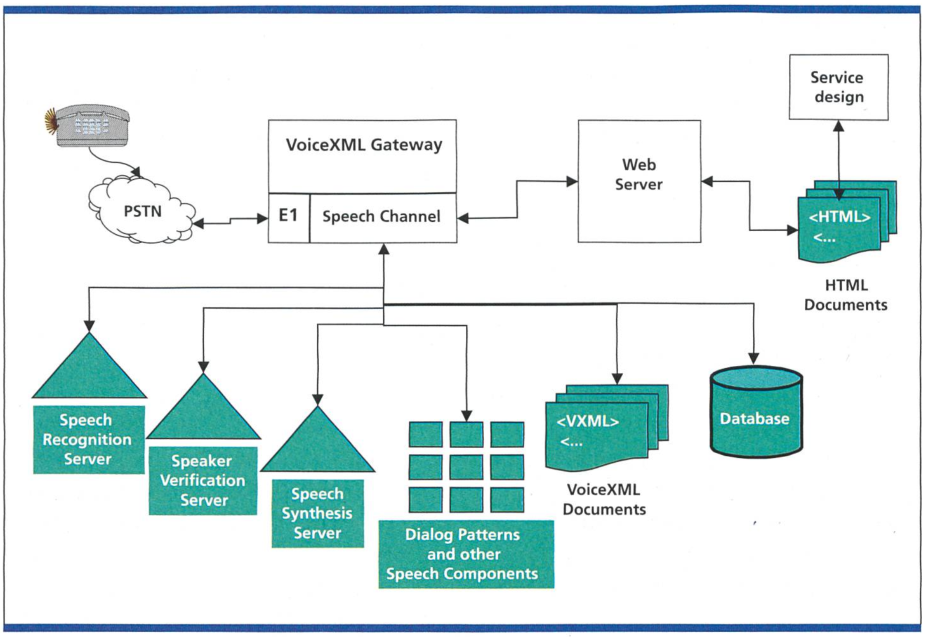

Fig. 2. The main components of the realised open voice services system contain a web server and a VoiceXML gateway. While the web server provides access to the service creation environment over HTML pages, the VoiceXML gateway controls the operative service components like the speech recognition server, the speech synthesis server, dialogue patterns and VoiceXML documents.

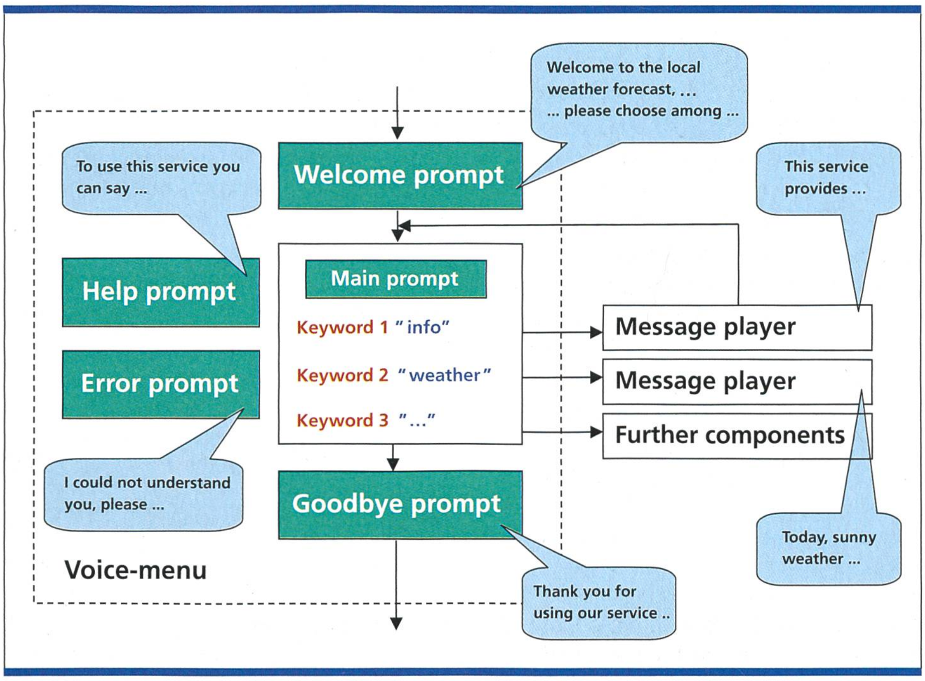

Fig. 3. The voice-menu dialogue pattern: The diagram shows the required service components and the dialogue flow. The messages show a possible example for a simple weather forecast application.

(for example another voice menu). At the end, a standardised help prompt, an error prompt and a goodbye prompt complete the voice menu (fig. 3).

Conclusions

Voice, as an easy-to-use man-machine interface, has become applicable in the last few years for certain application domains. Using a restricted vocabulary with a sound dialogue design will make these applications usable for a broad range of customers. However, at the moment the design of such speech-enabled services is rather time-consuming and therefore costly because dedicated expert know-how is necessary to setup the service.
By generalising the steps needed to build a speech-enabled service massive cost reductions can be achieved. Based on a unified webbased approach our demonstration system shows that it is possible to create, configure, and maintain a fully functional speech-enabled service without dedicated speech technology know-how, therefore accelerating and facilitating service creation significantly.
The current challenge within this technology is to anticipate dialogue patterns which are flexible and easy to use over a web interface. A trade-off between flexibility and complexity has to be made to fulfil the customer needs. With a representative list of voice patterns, each describing a specific service template, it should be possible to realise a large number of speech applications.
VoiceXML, interpreted by a VoiceXML gateway, is certainly one of the key success factors to realise the concept of open voice services.
The advantage of this technology is that the customer, that is the service provider, can create his own service – after all he knows best what he wants. Shorter time-to-market and lower prices are further benefits for customers and for Swisscom, respectively. Speech-enabled services become easy to use at a very low price in a wide range of possible application domains.

Outlook

The results of the first trial are encouraging. Further dialogue patterns, which can be combined with the existing ones, will be explored. Moreover, access to databases and dynamic information (such as weather data) available on the web will be treated. With a long-term view of 2–5 years and keeping in mind that almost all people in the western hemisphere have a voice access, the goal is to make voice-enabled services a mass product available to a huge number of users. 10

References

[1] VoiceXML Forum, Voice eXtendable Markup Language, VoiceXML, Version 1.0, March 2000. http://www.voicexml.org/spec.html

[2] L. Rabiner and B.-H. Juang, Fundamentals of Speech Recognition, Prentice Hall, Inc., 1993, ISBN 0-13-015157-2

[3] D. Borchers, Mein Computer versteht mich nicht, Probleme mit der automatischen Spracherkennung, Neue Zürcher Zeitung, 2nd of February 2001, Nr. 27, page 78

[4] M. Zink, Zusammenspiel Mensch-Maschine-Schnittstellen in Sprachsystemen, c't magazin für computer technik 3/1999, page 133, Heisse Verlag

[5] T. Dutoit, An Introduction to Text-To-Speech Synthesis, Kluwer Academic Publishers, ISBN 0-7923-4498-7

[6] B. Balentine and D.P. Morgan, How to Build a Speech Recognition Application – A Style Guide for Telephony Dialogues, Enterprise Integration Group, ISBN 0-9671278-1-5

Abbreviations

HTML	Hyper Text Mark-up Language
TTS	Text-to-Speech, Speech Synthesis
XML	Extendable Mark-up Language
VoiceXML	Voice Extendable Mark-up Language, a Subset of XML
VXML	Synonym for VoiceXML

VoiceXML	http://www.voicexml.org
Speech Recognition	http://svr-www.eng.cam.ac.uk/comp.speech/index.html
TTS	http://tcts.fpms.ac.be/synthesis/introtts.html
Project SPES (Closed User Group access only)	http://ctep.swissptt.ch/ep31/projects/speech.htm

Zusammenfassung

Gesprochene Sprache ist für den Menschen die einfachste und natürlichste Art Informationen auszutauschen, sei dies direkt oder über ein Medium wie beispielsweise das Telefon. Auf der anderen Seite hat sich das Internet in den letzten Jahren zu einer wichtigen Quelle des Informationsaustauschs etabliert. Die Vision der Kombination dieser beiden Kommunikations- und Informationskanäle mit der Absicht, deren Vorteile gegenseitig nutzbar zu machen, führt zu einer auf gesprochener Sprache basierenden Mensch-Maschine-Schnittstelle. Leider ist die Entwicklung von sprachgesteuerten Diensten eine sehr kosten- und zeitintensive Angelegenheit. In unserem Projekt sind wir bestrebt, dem Kunden einfache Dialogmuster über Web-Seiten zur Verfügung zu stellen. Mit diesen Mustern soll er in der Lage sein, rasch und unkompliziert seinen eigenen sprachgesteuerten Dienst aufzubauen.

Urs-Viktor Marti *received his MS and PhD degrees in computer science from the University of Bern, Switzerland, in 1996 and 2000 respectively. His doctoral work was on off-line hand-written sentence recognition. His research interests include neural networks, document image analysis, language modelling and speech recognition. Since March 2001 he is working for Swisscom AG, Corporate Technology as an engineer in the domain of speech recognition.*

Robert van Kommer *is working for Swisscom AG, Corporate Technology. Since 1992, he has initiated, together with research partners, several projects in the field of voice-activated services. Specifically, he led the Swiss Polyphone project that consisted of collecting the speech corpus necessary for deploying speech recognition in Switzerland. He is currently pursuing a PhD thesis in the PAI group of Prof. Beat Hirsbrunner at the University of Fribourg.*

Oliver Krone *studied computer science and electrical engineering at the Technical University of Munich and later received a doctoral degree from the University of Fribourg. After graduating from Munich, he worked as a Research Fellow at the IBM European Networking Centre in Heidelberg, Germany, where he participated in the development of a multimedia communications system. He joined Swisscom Corporate Technology in 1998 and is currently Manager of the Open Communication Services Architecture Exploration Programme.*

Muss man alle Technologiesprünge mitmachen?

Diese Frage stellen sich zunehmend Firmen im Zusammenhang mit der Halbleitertechnik. Übertragungsraten von 10 Gbit/s stehen in Zukunft an und die Technik für 40 Gbit/s ist bereits in der Vorentwicklung. Jetzt hat einer der Spezialisten in der amerikanischen Halbleiterindustrie eine klare Aussage gemacht: Die Broadcom Corporation hat erkennen lassen, dass man mit Standard-CMOS-Technologie und Strukturbreiten von 0,13 µm die meisten Anforderungen für 10 Gbit/s erfüllen kann. Nicht genug: Selbst für einige 40-Gbit/s-Anwendungen dürften Strukturbreiten von 130 nm auf den Chips noch ausreichen.

Broadcom Corp.
16125 Alto Pkwy
Irvine
CA 92618
USA
Tel. +1-949-450 8700
E-Mail: info@broadcom.com

Rambus verliert Prozess gegen Infineon

Ein amerikanisches Gericht hat die Rambus Corporation wegen Betrugs zu 3,5 Mio. US-$ Strafe verurteilt. Dem amerikanischen Unternehmen wurde vorgeworfen, es habe eigene Patente einem internationalen Gremium (JEDEC) absichtlich vorenthalten, um dann hinterher Klagen wegen Patentverletzungen einzureichen. Das Gericht stellte fest, dass weder in SDRAM noch in DDR-SDRAM von Infineon Patente von Rambus verwendet worden sind.

Infineon Technologies AG
Postfach 80 09 49
D-81609 München
Tel. +49-89-234 0

Was kommt nach dem 40-GHz-Band?

Die Vorarbeiten für die drahtlose Kommunikation im 10-GHz-Band sind weiterhin am Laufen und für das 40-GHz-Band liegen die ersten Grundlagenarbeiten vor. Doch bereits spricht man über das nächste Frequenzband. Die amerikanische Telekom-Aufsichtsbehörde FCC hat von der WCA (Wireless Communications Association) eine Anfrage bekommen, ob nicht das 94-GHz-Band langfristig für kommerzielle Nutzung freigegeben werden könnte. Hintergrund für das Interesse ist der Wunsch, mit dem Ethernet-Standard auch bei grösseren Datenströmen im Mobilfunkbereich Schritt halten zu können. Und die lassen sich leichter bei 100 GHz als bei 10 GHz bewältigen.

Ein optischer Router der Terabit-Klasse

Spektakuläres Schaustück auf der «Supercomm 2001» in Atlanta (Georgia) war die Vorführung eines photonischen MPLS-Routers (Multi-Protocol Lambda Switching) durch die japanische NTT Communications. Er steuert dynamisch sowohl das Routing in optoelektronischen Netzen als auch die belegte Bandbreite und erhöht damit kräftig den Datendurchsatz im Netz. Die Japaner sind die ersten, die einen solchen optischen Router für den Terabitbereich weltweit demonstrieren können. Das Unternehmen will diesen Superrouter zunächst noch im firmeneigenen Kommunikationsnetz einsetzen. Man erhofft sich davon nicht nur grössere Übertragungsbandbreiten im eigenen Netz, sondern auch deutlich verbesserte Zuverlässigkeit.

NTT Communications Corp.
1-1-6 Uchisaiwaicho
Chiyoda-ku
Tokyo 106-8019
Japan

«Drahtloses Internet» auf nur einem Chip

Noch ist es ein experimentelles Muster, was Intel im Labor fertiggestellt hat: Einen Chip, der in künftigen Mobilfunksystemen den kompletten Internetzugang übernehmen kann. Er enthält alle notwendigen Logikfunktionen, hat einen eingebetteten Flash-Speicher mit hoher Packungsdichte und die analogen Systembestandteile für die Netzkommunikation. Dieser Mixed-Signal-Chip fällt durch eine hohe Übertragungsleistung auf. Sichere Planungsdaten sollen erst zu einem späteren Zeitpunkt mitgeteilt werden.

Intel Corp.
3065 Bowers Avenue
Santa Clara
CA 95051-8126
USA
Tel. +1-408-765 8080