



Sprich mit mir!

Sprachausgabe als Bestandteil moderner Mensch-Maschine-Schnittstellen

Der Erfolg von mobilen und ab Werk verbauten Navigationslösungen geht einher mit einer rasanten Entwicklung der Bedienkonzepte. Ein Baustein ist dabei die Sprachausgabe, die neben der Zielführung auch zur Vermittlung weiterer Informationen genutzt werden kann, etwa des Vorlesen von Kurznachrichten oder E-Mails. Im Interview erklärt Dr. Martin Reber, bei der SVOX AG für Kundenprojekte zuständig, die Rolle des Sprachdialogdesigns innerhalb der Entwicklung von Mensch-Maschine-Schnittstellen. (Bild: PixelQuelle.de)

■ Herr Dr. Reber, welchen Stellenwert hat die Sprachausgabe bei Ihnen?

Dr. Reber: Sprachausgabe ist die Schlüsselkompetenz der SVOX AG. Allerdings bieten wir mit unserem SVOX-ExpertSpeech-System eine universelle Sprachausgabe, die frei konfigurierbar sowohl aufgezeichnete Textbausteine ausgeben kann als auch Text-to-Speech (TTS) beherrscht. Sprachausgabe ist dabei in Form eines Navigationssystems in praktisch jedem Fahrzeug enthalten.

■ Welche Rolle spielt das Sprachdialogdesign bei der Entwicklung heutiger und zukünftiger Automobile?

Dr. Reber: Eine heutige Head-Unit ist ein sehr komplexes System mit begrenzter Rechenleistung. Die Sprachausgabe soll auch dann nicht ins Stocken geraten, wenn beispielsweise im Hintergrund eine neue Route berechnet wird und gleichzeitig auf die Festplatte oder ein DVD-Laufwerk zugegriffen wird. Unsere Sprachausgabe benötigt je nach Konfiguration zwischen 20 und 30 MIPS Rechenleistung, was in etwa der Hälfte der Vorgaben der Hersteller entspricht. Möglich wird dies nur durch eine intelligente Software-Architektur (**Bild**) und einer optimal konfigurierbaren TTS, das heißt, die Anwendung liefert lediglich ein Stück Text, welches wir in Sprache überführen und mit aufgezeichneten

Ansageteilen kombinieren. Positiver Nebeneffekt dieser Architektur ist, dass sich unsere Software vergleichsweise einfach in bestehende Systeme integrieren lässt, in der Regel innerhalb einer Woche.

In Zukunft wird die Wichtigkeit des Dialogdesigns an sich noch weiter zu-



■ Dr. Martin Reber, Director Embedded Solutions und Mitglied der Geschäftsführung der SVOX AG: „Wir machen das Leben unserer Kunden sehr einfach.“

(Bild: SVOX AG)

nehmen. Zum einen finden sich bereits heute neue Funktionen wie etwa SMS-Reader und zum anderen müssen in der nächsten Generation sprachliche, visuelle und haptische Kanäle verbunden werden. Das Schlagwort dazu heißt Multimodalität.

■ Wie wichtig ist „Voice Branding“ für die Automobilhersteller?

Dr. Reber: Voice Branding, also das Bereitstellen von exklusiven Stimmen, ist bei den OEMs ein zentrales Thema, um den Wiedererkennungswert zu erhöhen. Speziell im Premiumsegment müssen die Stimme und auch das „Wording“ eine bestimmte Atmosphäre erzeugen, die den High-End-Anspruch erfüllt.

Wichtig ist aber auch, dass die Sprachausgabe richtig funktioniert und in mehrdeutigen Situationen korrekt reagiert. Beispielsweise kann „9 A“ einmal die „Autobahnausfahrt 9 A“ bezeichnen, in einer empfangenen E-Mail aber „9 Ampere“ bedeuten – und soll dann auch so ausgegeben werden.

■ Wie beurteilen Sie die Konvergenz von Geräten aus dem Bereich der Konsumelektronik und ins Fahrzeug integrierten Lösungen?

Dr. Reber: Gibt es diese Konvergenz wirklich? Ich sehe eher eine Spezialisierung der einzelnen Geräte, etwa spezielle Navigationslösungen für Autos, Fahrräder oder zum Wandern. Integrierte beziehungsweise fest eingebaute Systeme sind problematisch hinsichtlich Wartung und Updates, außerdem sind deren Eigenschaften aufgrund des längeren Produktlebenszyklus oft denen der mobilen Geräte nicht gewachsen. Beide Produktgruppen werden weiter existieren, sich gegenseitig sicher beeinflussen oder voneinander

lernen. Wir versuchen dabei, die Entwickler von fest eingebauten Systemen durch die Architektur des SVOX-ExpertSpeech-Systems zu unterstützen. Diese hilft, unnötige Schleifen bei Implementierung oder der Integration mehrerer Sprachen zu vermeiden.

■ **Mit welchen speziellen Eigenschaften können die neuen Geräte aufwarten?**

Dr. Reber: Am wichtigsten ist hier sicherlich das Thema der Cross-Lingualität. Eine deutsche SVOX-Stimme kann in Frankreich französische Städte mit deutscher Betonung aussprechen. Dieses „Signpost-Feature“, das auch in Low-end-Geräten zu finden sein wird, nutzt die phonetischen Datenbanken der Kartenhersteller des jeweiligen Landes. Die Herausforderung dabei ist, die ursprüngliche Stimme beizubehalten. Dadurch, dass wir für fast alle existierenden Lautschriften ein entsprechendes Interface bereitstellen, machen wir das Leben für unsere Kunden sehr einfach.

■ **An welchen Problemen beziehungsweise deren Lösungen wird derzeit gearbeitet?**

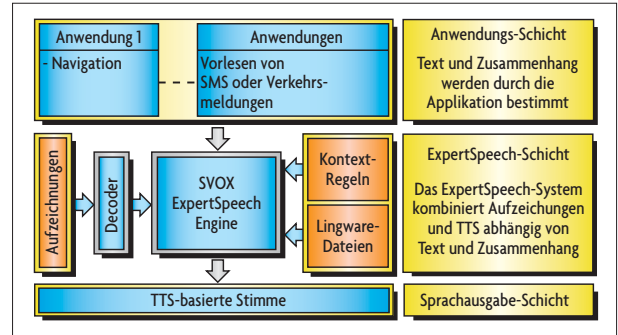
Dr. Reber: Auch hier steht die Mehrsprachlichkeit ganz vorne auf unserer Liste. In vielen Fällen ist die korrekte Aussprache ohne spezifisches Know-how sehr schwierig. Stellen Sie sich beispielsweise ein deutsches System vor, bei dem ein französischer Interpret angesagt werden soll, der englische Lieder singt. Mit der älteren Methode, aufgenommene Wave-Dateien abzuspielen, kann das nicht bewerkstelligt werden. Das SVOX-TTS-System besteht deshalb aus einer Engine und zusätzlichen Lingware-Dateien. Die Engine ist ein universelles Stück Soft-

ware, das die entsprechenden Methoden und Algorithmen enthält, um Sprache zu erzeugen. Die Lingware-Dateien umfassen die charakteristischen Eigenschaften einer Sprache. Durch diese Trennung von Abläufen und Daten bietet sich die Möglichkeit, Kontext-Informationen einfließen zu lassen oder das System an neue Sprachen anzupassen. So gibt es entsprechende Lingware-Versionen für unterschiedliche Sprachen, etwa asiatische, osteuropäische oder aus dem EU-Raum, die aber alle mit der gleichen Engine funktionieren.

■ **Welche Vorteile bietet der Einsatz des SVOX-TTS-Systems?**

Dr. Reber: In der Architektur unserer Software wird ganz klar zwischen den Abläufen und den Datensätzen unterschieden. Durch diese zusätzliche Abstraktionsebene agieren wir praktisch als „Black Box“, bei der auf der einen Seite der Text eingespeist wird und auf der anderen Seite PCM-Samples ausgegeben werden. Will etwa der Tier-1-Zulieferer den Text von „Jetzt links fahren“ nach „Und jetzt bitte links fahren“ ändern, muss er lediglich in seiner Entwicklungsumgebung ein Textupdate einspielen, bei uns müssen ein paar Konfigurationsdateien angepasst werden. Bisher musste man für derartige Änderungen hunderte von Text-Bruchstücken durchforsten und abändern. Bei zwei bis drei Updates pro Entwicklung und teilweise bis zu 15 Sprachen pro System ist die Zeiterparnis gravierend, die wir hier bieten.

Ein weiterer Vorteil ist natürlich die Simulation. Sowohl unsere Engine als auch die Konfigurationsdaten sind plattformunabhängig. Die Software läuft auf einem normalen PC



und wird nach Bedarf auf die Zielhardware portiert. So können unsere Kunden frühzeitig erfahren, wie das System klingt.

■ **Im Hinblick auf die Hersteller von entsprechenden Systemen: Was sehen Sie im Bereich der Dialogsysteme als deren größte Herausforderung?**

Dr. Reber: Eines der größten Themen ist sicher die bereits angesprochene Multimodalität. Auch wenn die Zieleingabe per Sprache getätigt wird, muss die Navigation zunächst per Knopfdruck aktiviert werden. Die Zielbestätigung erfolgt dann grafisch auf einem Display, bei Mehrdeutigkeiten erfolgt eine manuelle Auswahl per Knopf. Der gesprochene Befehl „Start“ lässt dann die Zielführung beginnen. Dieser ständige Wechsel zwischen Haptik und Spracheingabe und auch die Konsistenz zwischen visuell dargestellten und über Sprachausgabe gelieferten Daten bietet noch viel Entwicklungspotential, sowohl für die Zulieferer als auch für die OEMs. Wir liefern zwar entsprechende Simulationssoftware für die Entwicklung, aber die eigentliche Gestaltung der Dialoge liegt in der Verantwortung der Hersteller.

Das Interview führte Stephan Janouch von der „Elektronik automotive“.

■ **Das SVOX-ExpertSpeech-System basiert auf einer zentralen Engine. Sprachbausteine, Text für die TTS-Umwandlung oder Konfigurationsinformationen erhält diese über definierte Schnittstellen.**

(Quelle: SVOX AG)