

Oktober 07

Auf den Spuren von „Sandra“ – oder wie kommt die Stimme ins Navi?

Sprachansagen bei Navigationssystemen sind heute selbstverständlich. Hier erfahren Sie, wie dieses Feature erstellt wird – von Aufnahmen im Tonstudio über Text-to-Speech bis hin zur Übersetzung der Texte in verschiedene Sprachen.

„Bitte in zweihundert Metern rechts abbiegen!“ fordert eine angenehme Frauenstimme Volker Hartmann, Senior Product Manager bei der PTV AG, auf. Auf die „sanfte Sandra“ kann er sich immer verlassen – sie führt ihn beim Autofahren stets sicher ans Ziel. Dabei beschränkt Sandra sich auf die reine Wegansage, denn die Stimme der „klugen Beifahrerin“ kommt aus dem Navigationsgerät. Hartmann, der bei der PTV im Bereich mobile Navigation tätig ist, weiß bestens darüber Bescheid, wie die Sprachansagen im Tonstudio produziert werden. Daneben gibt es heute auch rein synthetische Computerstimmen und eine Kombination aus menschlicher und digitaler Stimme, das so genannte Text-To-Speech. Hartmann gibt einen Einblick in die Welt der Sprachansagen für die mobile Navigation.

Für Navigationssoftware mit menschlichen Sprachansagen werden heute Originalstimmen im Tonstudio aufgenommen, die nur die allgemeinen Textteile sprechen. Orts- und Straßennamen erscheinen ausschließlich auf dem Bildschirm. „Diese zusätzlich zu integrieren, würde einen riesigen Speicherplatz benötigen“, weiß Hartmann. „Ein Stimmensatz, also alle Textteile die zusammen eine korrekte Sprachansage ergeben, hat eine Größe von etwa vier MB.“

Die klangvollste Stimme muss nicht die beste sein

Voreingestellt bei deutschsprachigen Navigationsansagen ist die weibliche Stimme „Sandra“. Welche reale Person hinter der beliebtesten Sprachansage bei der Navigationssoftware von PTV steckt, verrät Harald Elsen, Geschäftsführer des Übersetzungsbüros Delta International CITS GmbH in der Holtorfer Straße in Bonn; Aus dem Tonstudio dieser Agentur stammen die Aufnahmen für die Fahrhinweise des Navis: „Die Stimme gehört der Germanistin Petra Niefind, die bei einer Sprachschule Deutsch als Fremdsprache unterrichtet und für mein Übersetzungsbüro als freiberufliche Lektorin arbeitet.“ Niefinds Stimme eignet sich aus mehreren Gründen. Sie spricht akzentfrei und hat den Umgang mit Sprache im Rahmen von Stimmbildungs-Kursen erlernt. Hartmann meint: „Die klangvollste Stimme ist nicht unbedingt am besten geeignet. Ein Kriterium ist die Verständlichkeit. Auch im Straßenverkehr mit vielen

Neben- und Störgeräuschen muss die Sprachansage noch gut zu verstehen sein. Nicht alles, was im Büro beim Test gut klingt, eignet sich für die tatsächliche Benutzung im Auto.“

Elsen, der die Tonaufnahmen in seinem Hause produziert, nennt weitere Kriterien bei der Stimmenwahl: „Die Einzelteile müssen so eingesprochen werden, dass Betonung und Satzmelodie beim späteren Zusammensetzen stimmig sind. Deshalb sollten die Sprecher gut mit ihrer Stimme umgehen können und ein hohes Maß an Artikulationsfähigkeit besitzen.“ Häufig haben die Sprecher deshalb eine musikalische Vorbildung.

Wer diesen Anforderungen entspricht, kann sich bei Delta als Sprecher bewerben: „Auf unserer Website und an Unis schreiben wir aus, wenn wir nach Sprechern suchen. Meist benötigen wir Muttersprachler aus verschiedenen Ländern für Sprachen von Afrikaans über Russisch bis Urdu“, merkt Elsen an.

Große Studios, die beispielsweise auch Radio- und Werbebeiträge aufnehmen, arbeiten häufig mit professionellen Sprechern zusammen, die bereits Erfahrungen mit Werbespots oder beim Synchronisieren gemacht haben.

So funktioniert die Produktion von Sprachansagen

Für eine Aufnahme schickt der Produzent eine Auswahl von zwei bis zehn unterschiedlichen Probeaufnahmen verschiedener Sprecher an den Navigationssoftwarehersteller, woraus dieser nach Beratung mit Muttersprachlern seinen Favoriten auswählt.

Der Sprecher im Studio bekommt eine Liste mit allen benötigten Textteilen, aus denen sich die Ansagen später zusammensetzen. Diese spricht er dann hintereinander ein; das dauert etwa eine Stunde. Niefind, die auch heute noch des Öfteren für Sprachaufnahmen ins Studio geht, beispielsweise für E-Learning-Programme, beschreibt, wie die Aufnahmen bei Delta vor zwei Jahren abliefen: „Im Studio ist eine kleinere schalldichte Box eingebaut – darin ist selbst das Schlucken ungewöhnlich laut! Ich stand an einer Art Notenständer mit den Texten und habe alle Textteile nacheinander eingelesen, während ein Techniker draußen alles aufgenommen und überwacht hat.“

Für die Nachbearbeitung sind mehrere Arbeitsschritte nötig: Erst wird kontrolliert, ob alle Textmodule korrekt aufgenommen wurden. Dann sind alle Einzelteile als wav-Dateien abzuspeichern, wobei beachtet werden muss, dass sie richtig benannt werden. Nur so ist sichergestellt, dass die Dateien später problemlos in der richtigen Reihenfolge aufgerufen werden können. Zudem darf die Lautstärke der Aufnahmen nicht über sechs Dezibel hinausgehen, damit sie später im Navi verwendbar sind.

Produziert wird zunächst eine PC-Version der Software mit der Sprachansage. Diese ruft in der richtigen Reihenfolge die Einzelteile auf, z. B. „Biegen Sie“, „in zweihundert Metern“, „rechts“ und „ab“.

„Nur wenn alle Aufnahmen korrekt sind, kann beim Test mit der vorläufigen PC-Software live auf der Straße auch alles funktionieren, so dass die Testpersonen ihr Ziel erreichen“, kommentiert Hartmann. „Manchmal klappt es auf Anhieb, in anderen Fällen muss noch viel Aufwand in Korrekturen gesteckt werden. Wir legen großen Wert auf diese Tests, da wir nur dadurch eine hohe Qualität der Sprachansagen garantieren können.“

Sprachansagen in 28 Sprachen

Der Softwarehersteller PTV verfügt im Deutschen über zwei weibliche Stimmen, die sanfte „Sandra“ und die dominantere „Angela“, sowie über die männliche Stimme „Thomas“. Daneben sind 27 weitere Sprachen – in der Regel eine weibliche und eine männliche Stimme – vorhanden; zwei weitere, Malaiisch und Mandarin, sind in Arbeit. Dabei sind beispielsweise im Englischen verschiedene Sprachvarianten verfügbar, derzeit für Australien, das Vereinigte Königreich und die Vereinigten Staaten.

Neben der Prosodie, das heißt der korrekten Akzentuierung der Worte, spielt bei der Produktion der Sprachensätze die jeweilige Syntax, also die Zusammenstellung der Satzbausteine, eine tragende Rolle. Je nach Syntax müssen die Bausteine in einer anderen Reihenfolge kombiniert werden als die deutschen Vorlagen. In manchen Fällen ist es sogar erforderlich, einzelne Module zu splitten.

„Die meisten indo-europäischen Sprachen lassen sich recht gut umsetzen“, erklärt Marie Thiriet, Leiterin der Übersetzungsabteilung bei PTV. „Probleme gibt es beispielsweise bei agglutinierenden Sprachen, wie Finnisch, Türkisch und Ungarisch, da dabei Elemente, so genannte Affixe, an die Wörter angehängt werden.“

Verschiedene Karten- und Sprachensätze in Kombination

Normalerweise erfolgen Sprachansage und die synchrone Anzeige auf dem Bildschirm in derselben Sprache. Allerdings werden bei einigen Sprachen mit anderen Schriftzeichen als die lateinischen Buchstaben, zum Beispiel bei Arabisch oder Chinesisch, englische Texte angezeigt.

Möglich sind aber auch jeweils verschiedene Karten- und Sprachensätze in Kombination. „Unsere Kunden erhalten von uns als B2B-Anbieter unterschiedliche Kombinationen für ihre Geräte. Zielt ihr Angebot beispielsweise auf den potenziellen Kundenkreis „türkische Autofahrer in Deutschland“ ab, dann wird eine Deutschlandkarte mit türkischer Sprachansage aufgespielt“, erläutert Hartmann.

Der Ex-Kanzler sagt, wo´s langgeht

Einige Navigationsanbieter bieten Fun-Stimmen zum Herunterladen auf das Navigationsgerät an. Beispielsweise führt der „Ex-Kanzler“ oder „Der Kaiser“ den Autofahrer sicher ans Ziel. Diese Aufnahmen stammen in der Regel von Stimmimitatoren.

Des Weiteren gibt es Ansagen in Dialekten, beispielsweise in Bayrisch. „Promi- und andere Fun-Stimmen sind vor allem beim jüngeren Publikum beliebt. Wir bieten solche Zusatzstimmen nicht an, da unser Zielpublikum bei der mobilen Navigation in der Altersklasse von Mitte 30 bis 60 Jahren liegt“, begründet Hartmann.

Navigieren ganz ohne Bildschirm mit Text-To-Speech

Neben Originalstimmen werden heute synthetische, das heißt digital erzeugte, Stimmen verwendet, so genannte Text-To-Speech-Sprachansagen (TTS). Dafür wird aus Phonemen einer Originalstimme eine synthetische Stimme entwickelt, welche die Texte als Ton ausgibt. Das ermöglicht eine detaillierte Ansage der Fahrhinweise, bei der auch Straßen- und Ortsnamen vorgelesen werden. Beispielsweise sagt die Stimme nicht nur „Bitte in zweihundert Metern rechts abbiegen“, sondern „Bitte in zweihundert Metern rechts in die Blumenstraße abbiegen.“

Außerdem werden Signpost-Informationen, also reale Beschilderungsanzeigen zur leichteren Navigation ausgegeben. Verkehrsmeldungen können mit Hilfe von TTS so gesprochen werden, dass der Blick auf den Bildschirm überflüssig wird. Eine Umfrage der Universität Bonn in Zusammenarbeit mit pocketnavigation.de von 2005 hat ergeben, dass Autofahrer besonders Wert auf präzise Angaben zur Fahrtrichtung sowie auf die Nummerierung, Straßen und Ortsnamen legen.

„Das Ziel von TTS bei der Navigation ist, dass der Fahrer den Bildschirm nicht mehr benötigt, sondern sich voll und ganz auf den Straßenverkehr konzentrieren kann“, erklärt Hartmann.

Die beste Qualität wird durch eine Kombination erreicht: Eine Originalstimme spricht die allgemeinen Satzteile, Detailangaben, beispielsweise Straßennamen, werden mit Hilfe der synthetischen Stimme eingeblendet. Für diese künstlichen Tonausgaben wird die Stimme des Sprechers digitalisiert und so die synthetische Stimme erzeugt. Bei der kombinierten TTS-Ansage müssen von der Originalstimme gesprochene Satzteile und digitalisierte Ansagen sowohl in der Syntax als auch in der Wortbetonung so aufeinander abgestimmt sein, dass der Hörer nicht mehr merkt, dass es sich um eine kombinierte Ansage handelt.

Schwer aussprechbare Ortsnamen lassen sich mit der synthetischen Stimme dann schwer nachbilden, wenn diese spezielle Laute enthalten. Soll beispielsweise eine synthetische Stimme, die auf Grundlage einer deutschen Originalstimme erstellt worden ist, französische Namen aussprechen, die Nasale beinhalten, ist die Aussprache oft undeutlich. „Da gibt es erhebliche Qualitätsunterschiede. Die synthetischen Stimmen mancher TTS-Anbieter eignen sich beispielsweise gut zum Nachbilden der romanischen Sprachen, während andere wiederum englische Laute gut ausgeben“, erzählt Hartmann.

Rein synthetische Stimmen für die Onboard-Navigation

Rein synthetische Stimmen werden heute noch bei der Routenplanung am PC verwendet, bei der so genannten Onboard-Navigation. Dabei wird der vom Benutzer vor Fahrtantritt vorgeplante Fahrtweg einfach hintereinander weg gesprochen. Berücksichtigt wird nicht, ob der Autofahrer sich tatsächlich noch auf der vorgeplanten Wegstrecke befindet oder ob es einen alternativen Weg gewählt hat. Dagegen erfolgt bei der mobilen Navigation eine Neuberechnung der Route via Satellit und aktuelle Daten werden eingespielt.

Sprecherin Niefind schlägt am liebsten im herkömmlichen Auto-Straßenatlas nach. „Auch ohne Navi finde ich mich meist gut im Straßenverkehr zurecht. In einer verkehrsreichen Großstadt im Ausland, zum Beispiel in Tokio, wäre ich allerdings dankbar, wenn mich eine Stimme in meiner Muttersprache durch das Verkehrschaos leitet.“ Ihre eigene Stimme hat sie im Auto noch nie gehört. „Aber wahrscheinlich würde ich meine eigene Stimme selbst gar nicht wieder erkennen“, schmunzelt sie. Hartmann dagegen lässt sich gerne von „Sandra“ nach Hause geleiten und kommt mit ihrer Hilfe jederzeit schnell und sicher am gewünschten Ziel an.

Bildmaterial



Mit dem mobilen Navigationsgerät schnell und sicher ans Ziel.

Weiterführende Informationen

- ▷ Umfrage der Universität Bonn zu Sprachausgabe:
http://www.pocketnavigation.de/article/view_496/1.1.0.html
- ▷ Internetseite von Delta International CITS GmbH:
http://www.dicits.com/index_1280.html
- ▷ Internetseite zu mobiler Navigation bei PTV:
http://www.ptv.de/cgi-bin/mobility/mob_oem.pl