

# VOICE CONTROL - READY FOR AUTOMATION

Warum Sprachinteraktion  
in Zukunft grafische  
Bedienkonzepte erweitern  
oder sogar ablösen wird





## **Industrielle Sprachsteuerung: Embedded, robust, natürlichsprachig**

Mit der "Spectra PowerBox 100-Voice" stellen die Spectra GmbH & Co. KG (Spectra) und der Sprachexperte voice INTER connect GmbH (VIC) ein erstes Ergebnis ihrer intensiven Zusammenarbeit auf dem Gebiet industrieller Sprachsteuerungen vor.

„Das gemeinsame Produkt erwächst aus unserem Glauben an den erheblichen Mehrwert von Sprachbedienungen in Industrie-Anwendungen“ erläutert Geschäftsführerin Diane Hirschfeld.

„voice INTER connect bringt Expertise und das spezielle Know-how im Bereich der Spracherkennung, -übertragung und -verarbeitung, sowie langjährige Erfahrungen bei der Entwicklung praxistauglicher Anwendungen ein. Die Spectra ist Branchenspezialist für Industrie-PC sowie Steuerungs- und Automatisierungstechnik.“

Die Interaktion mit Maschinen via Sprache ermöglicht als neue Modalität eine bisher nicht gekannte Effizienz und Ergonomie von HMI-Lösungen und wird in ausgewählten Bereichen herkömmliche grafische Bedienkonzepte anreichern oder ganz ablösen.

Maschinen werden als technische Assistenten den Menschen durch kontextbezogene Erfassung und Bereitstellung der richtigen Informationen in Echtzeit bei Entscheidungsprozessen unterstützen, sowie die Dokumentation und Qualitätssicherung von Produktionsprozessen rationalisieren.

Diese Trends greifen wir auf und bieten eine leistungsstarke Embedded-Lösung für die Industrie, die auch im lauten Umfeld und ohne Internet aus natürlich gesprochenen Sätzen eindeutige Befehle an die Maschine erkennt. Die Systemantworten oder benötigte Informationen werden via Sprache wieder ausgegeben.



## Welche Vorteile bietet Sprachbedienung?

Anders als im Smart Home oder Consumer-Umfeld, wo Sprachbedienung oft nur ein Komfortmerkmal ist, bietet Sprachbedienung in Industrieanwendungen echte Vorteile:

- Wenn Hände oder die Aufmerksamkeit des Anwenders durch manuelle Tätigkeiten wie Montage, Transport, die Handhabung von Medizintechnik oder Laborausüstung bereits gebunden sind, kann Sprache als „dritte Hand“ die Bedienung von Maschinen oder die Abfrage von Informationen dennoch parallel ermöglichen.
- Unter widrigen Umweltbedingungen, im Dunkeln, bei Hitze oder Kälte bzw. im sterilen Umfeld ermöglicht Sprache eine berührungslose Auslösung von Funktionen.
- Parallel zu manuellen Aufgaben erlaubt Sprache die Dokumentation von Arbeiten, z.B. im Pflege- und Servicebereich, oder die Beurteilung von Maschinen, wie im Gutachterwesen.

- Der Zugriff auf Funktionen per Sprache erfolgt direkt, ohne Umwege über Menüs wie bei GUIs. Weiterhin werden Ähnlichkeiten bei den Sprachbefehlen ausgewertet, so dass diese nicht präzise eingegeben werden müssen. Beide Eigenschaften bewirken eine Beschleunigung des Bedienvorganges.
- Sprachbedienung erlaubt eine extreme Miniaturisierung des User Interfaces. Im einfachsten Fall besteht das User-Interface aus einem Mikrofon, einem Lautsprecher und einer Taste, so dass Sprachbedienung auch für leichtgewichtige mobile Geräte wie SmartWatches und SmartGlasses geeignet ist.

Wie diskutiert, kann im industriellen Umfeld Sprachinteraktion in vielen Bereichen seine Vorteile gegenüber bisher üblichen Bedienmodalitäten ausspielen. Wie wir in den nächsten Abschnitten sehen werden, ist der Aufwand für eine Integration einer Sprachbedienfunktion aufgrund unseres modellbasierten Ansatzes gering, daher lohnt sich eine Bewertung des jeweiligen Nutzens.

## Produktvorteile der Spectra PowerBox Voice

Hochwertiger Industrie-PC

Web-Applikation für  
individuelle Sprachdialoge

Keine speziellen  
Vorkenntnisse nötig

Die Spectra PowerBox 100-Voice besteht aus einem Industrie-PC mit Lautsprecher, Mikrofon, einer hochwertigen Soundkarte und einem vorinstallierten Software-Paket von vicCONTROL industrial.

Außerdem erhält der Käufer eine Lizenz zur kostenlosen Nutzung der Basisfunktionen der vicSDC Web-Applikation. Mit dieser können Sprachdialoge erstellt, getestet und validiert werden. Die kundenspezifischen Dialoge können abschließend via Download auf die eigene Applikation übertragen werden.

Bei der Produktgestaltung haben wir Wert darauf gelegt, dass das Produkt für Anwender geeignet ist, die keine speziellen Vorkenntnisse mit Sprachbedienung haben. Mit der Spectra PowerBox 100-Voice soll der Anwender ein Tool erhalten, um umgehend beginnen zu können.

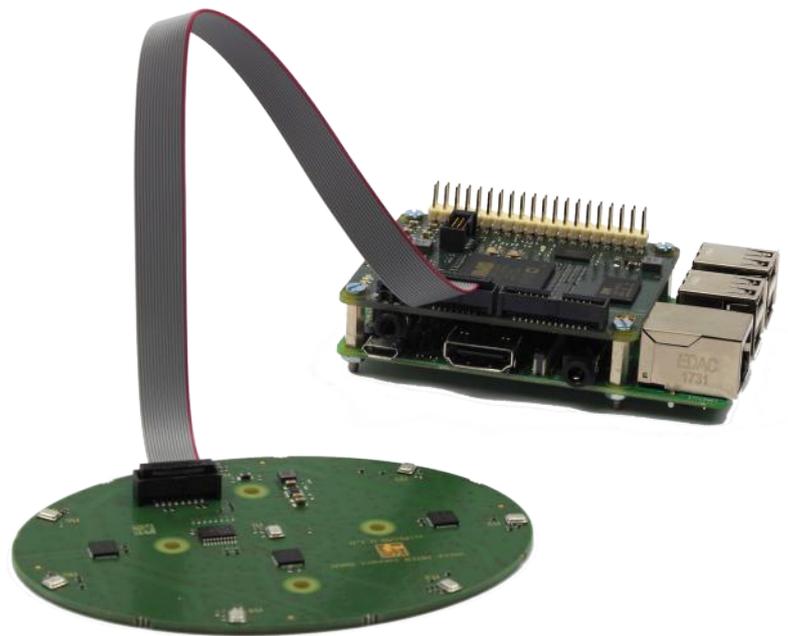


Anwendungen der Automatisierungs- und Fertigungstechnik erscheinen aufgrund der starken Umweltgeräusch-Belastung in Produktionshallen als wenig geeignet für eine Sprachbedienung.

Der herkömmliche Ansatz zur sauberen Spracherfassung ist die Ausstattung des Bedienpersonals mit geräuschfilternden und drahtlosen Headsets, welche an die Sprachverarbeitungseinheit angeschlossen sind.

Aufgrund der Bewegungseinschränkung, die das mit sich bringt, kann dieser Ansatz bei den Nutzern zu Akzeptanzeinbußen führen. Ebenso muss eine zuverlässige Funktion drahtloser Headsets für die Dauer einer Schicht gewährleistet sein bzw. das Ladekonzept muss sich gut in den Arbeitsablauf einfügen.

VIC bietet eine Lösung anderer Art. Durch spezielle Raumsensoren (vicDIVA-Technologie – Distant Voice Acquisition), die dreidimensional Sprachquellen verfolgen und deren Sprachsignale aufzeichnen, verstärken und von Umweltgeräuschen reinigen können, wird das Tragen körpernaher Sensoren (Headset,



Smartwatch) überflüssig. Die 3D-Audiosensoren haben je nach Art und Pegel der Umweltgeräusche eine Flächenabdeckung bis zu 70 m<sup>2</sup>, so dass bei entsprechender Installationsdichte auch größere Anlagen gut abgedeckt werden können. Auch für mobile Anwendungen wie autonome Transportsysteme oder Industrieroboter ist die Technologie aufgrund ihrer kompakten Bauweise und des geringen Stromverbrauchs gut geeignet. Mittels vicDIVA erreichen VICs Sprachbedienlösungen eine beispiellose Robustheit gegenüber Umweltgeräuschen.

## Kann Sprachbedienung in einer Industriehalle funktionieren?



## Gibt es Anwendungen, die sich nicht für eine Sprachsteuerung eignen?

Gerade im Bereich Automatisierungstechnik oder bei der Nachrüstung von Bestandsanlagen ist durch ein hohes Sicherheitsbedürfnis bzw. eine fehlende Netzwerk-anbindung eine lokale, integrierte Sprachbedienung wie vicCONTROL industrial gegenüber den bekannten, cloudbasierten Lösungen im Vorteil:

- Sie funktioniert in Echtzeit, ohne Latenzen durch die Sprachübertragung ins Internet.
- Sie ist ständig verfügbar, auch bei Ausfall vom Netzwerk oder sogar der Stromversorgung.
- Durch die lokale Verarbeitung treten auch keine Sicherheitslücken auf, die Daten bleiben beim Anwender.
- Durch den zugeschnittenen Wortschatz (domänenspezifischer Sprachdialog) wird eine hohe Verarbeitungssicherheit und Bedienungsgeschwindigkeit erreicht.

Den größten Nutzen erzielt der Anwender bei direkten Steueraufgaben, z.B. „Motordrehzahl 1.200 Umdrehungen einstellen“. Graduelle Regelprozesse hingegen („Licht heller.. heller.. stop“) lassen sich besser mit haptischen Bedienelementen wie Drehreglern abbilden.

Der Einsatz von Sprachbedienung darf ebenfalls nicht dazu führen, dass der Anwender mobile Maschinen unbeaufsichtigt oder versehentlich aktiviert. Laut Maschinenrichtlinie darf keine Sicherheits- oder Gesundheitsgefährdung des Menschen auftreten. Durch sorgfältige Beachtung prozessbezogener und ergonomischer Faktoren sowie Risikobetrachtungen zu Projektbeginn, können solche Gefährdungen bereits beim Entwurf der Lösung erkannt und vermieden werden. Darüber hinaus führt VIC bei den meisten Projekten eine Befragung und intensive Einbeziehung potentieller Nutzer- und Interessensgruppen im Rahmen von Feldtests durch, um alle Anforderungen beim Lösungsdesign berücksichtigen zu können, und bestmögliche Rationalisierungseffekte zu erzielen.

Safety-Anwendungen, bei denen allein eine Bedienung per Sprache stattfinden soll (z.B. Sprachbasiertes Not-Aus), müssen je nach Einsatzbedingungen beurteilt werden.

## Kundenspezifische Sprachlösungen - ein Kinderspiel?

Tatsächlich ist es ja so, dass der Mensch bei Gefahr intuitiv seine Stimme einsetzt, um sein Umfeld zu warnen oder sich bemerkbar zu machen. Dies kann man sich bei Notruf- und Notschalter-Anwendungen zunutze machen.

In Fällen mit lauten Hintergrundgeräuschen, z.B. in einer Maschinenhalle, ist eine Kombination mit komplementären Sensoren angebracht, da die Auswertung eines akustischen Sensors allein zu viele Risiken birgt. In Umfeldern wie Büros oder Wohnungen mit moderaten Hintergrundgeräuschen kann ein sprach-basierter Notruf durchaus komfortabler und sicherer als herkömmliche Anwendungen mit mechanischen Rufknöpfen sein, die im Notfall gegebenenfalls nicht am Körper getragen werden. In jedem Fall muss das Einsatzszenario analysiert und potentielle Risiken identifiziert werden, sowie eine nachhaltige Erprobung und Optimierung unter Einsatzbedingungen gegeben sein, damit die gewünschte Funktion im Ernstfall zuverlässig gewährleistet werden kann.

Bei Anwendungen, in denen die Bediener unterschiedliche Nationalitäten haben, oder eine Sprache mit Akzent sprechen, kann durch die Auswahl der passenden Zielsprache oder die nachträgliche Adaption des Wortschatzes bzw. Berücksichtigung von Aussprachevarianten, im Rahmen einer Felderprobungsphase die Funktionssicherheit optimiert werden.



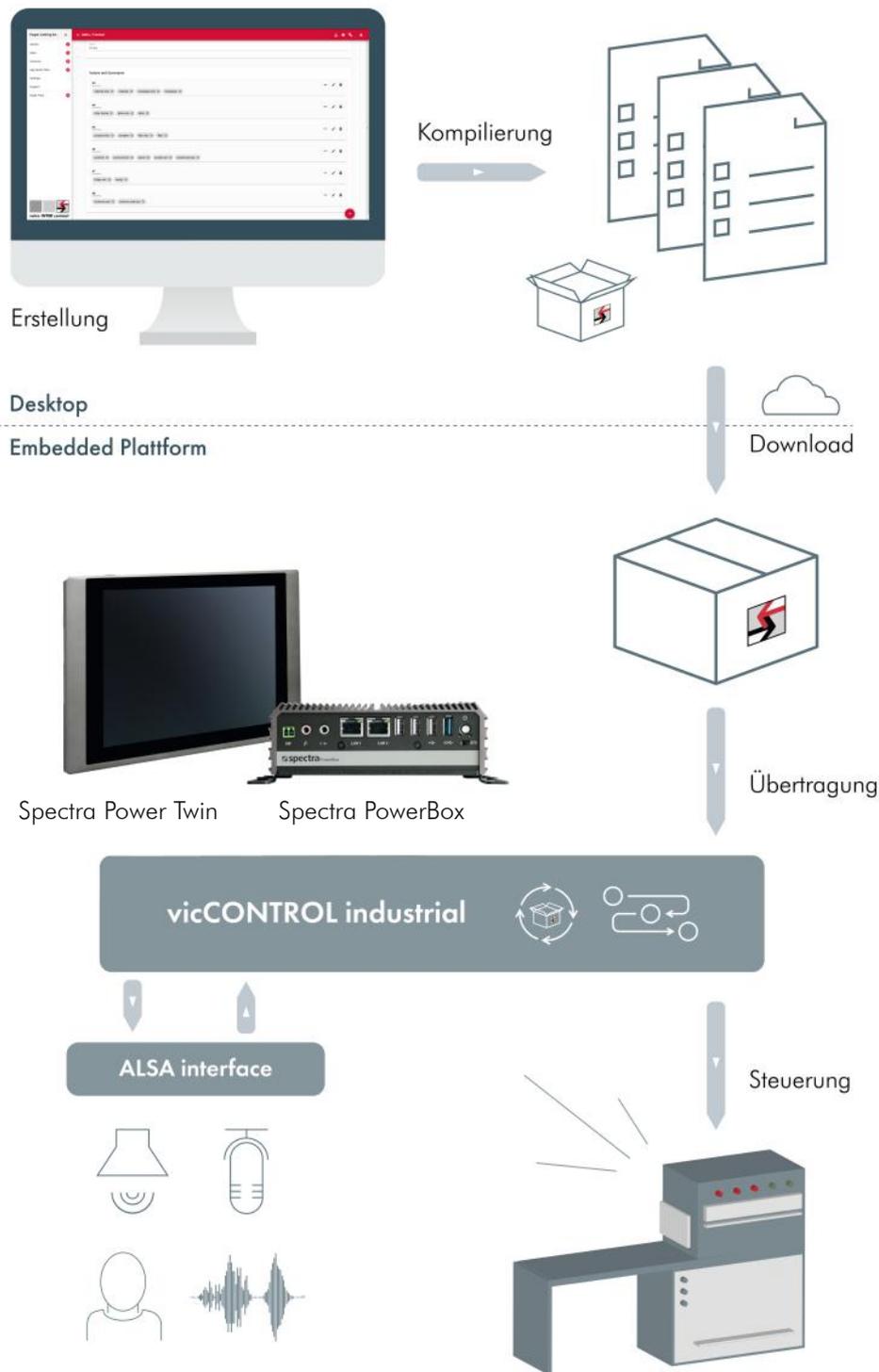
Mit den richtigen Werkzeugen und einer universellen, einfach zu bedienenden Schnittstelle zur Zielapplikation ist die Erweiterung vorhandener Maschinen und Anlagen um Sprachbedien-Funktionen ein Kinderspiel.

Aus ergonomischer Sicht ist zu beachten, dass eine komplette Substitution des herkömmlichen grafisch-haptischen Bedienkonzeptes (GUI – Graphic User Interface) durch ein Sprachbedienkonzept (SUI – Speech User Interface) für den Anwender eine anspruchsvolle Lernphase zur Folge hat. Günstiger für den Anwender ist es zunächst, das bekannte Bedienkonzept durch Sprachbedienung zu ergänzen (multimodale Bedienung). Dies lohnt sich umso mehr, wenn sich handfeste Vorteile aus der Sprachbedienung für den Anwender ergeben, wie zum Beispiel der Wegfall einer langwierigen Suche nach selten benutzten GUI-Bedienfunktionen. Nebenbei „lernt“ der Anwender das Steuervokabular durch geschickte Beschriftung der Bedienelemente in der GUI – es ergeben sich Synergien zwischen GUI und SUI.

Damit multimodale Bedienung dem Anwender zu jedem Zeitpunkt die Wahl seiner Vorzugsmodalität erlaubt, sollten alle Modalitäten synchron und verzögerungsfrei laufen.

In der Praxis erweist es sich als wirkungsvoll, wenn zwischen Anwendung und User Interface eine Abstraktionsschicht eingezeichnet wird (Broker), welche Steuer-Events und Zustände zwischen den Modalitäten und der Zielanwendung vermittelt.

Die Schnittstelle zur Zielapplikation soll einfach und ggf. durch Menschen lesbar, nachträglich erweiterbar und echtzeitfähig sein. Trotz sprachabhängiger Variationen einzelner Sprach-Befehle soll die Steuerschnittstelle eindeutig und sprachunabhängig sein.



Weiterhin sind Sprachbedienung und gesteuerte Applikation unter Umständen räumlich getrennt bzw. nur durch ein (IP-) Netzwerk verbunden (z.B. im Falle der Verwendung separater, intelligenter 3D Audiosensoren).

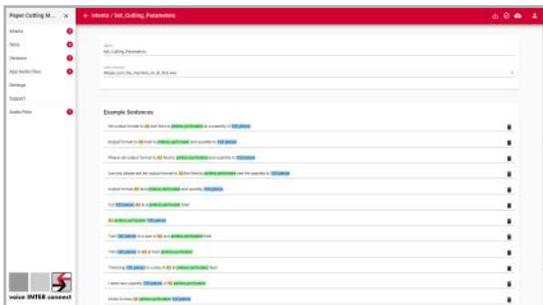
Für den Fall der Störung des Netzwerkes oder temporärer Überlastung ist ein Protokoll mit „Gedächtnis“, einer hohen Robustheit gegen Verbindungsabbrüche und einer gewissen Un-

abhängigkeit zwischen Sensor, Verarbeitungseinheit und Maschine wünschenswert. Das IoT Protokoll MQTT erfüllt all diese Anforderungen und wurde daher als Standardschnittstelle für das Sprachdialog-Interface ausgewählt. Die Protokollinhalte werden durch das vicSDC-Tool im Rahmen des Sprachdialog-Designs automatisch generiert, so dass für den Anwender kein zusätzlicher Integrationsaufwand anfällt.



## Erstellung kundenspezifischer Sprachbedien-Applikationen

Bei einem NLU-Erkennen (Natural Language Understanding) - wie der in vicCONTROL verwendete - können die Befehle und Parameter in beliebiger Reihenfolge und eingebettet in beliebige Trägerphrasen erkannt werden. Die Struktur eines Befehls ist also nicht an starre Regelwerke gebunden, sondern dieser kann ganz frei und komfortabel formuliert werden. Das Training eines kundenspezifischen Wortschatzes für einen solchen Erkennen ist deutlich einfacher geworden und kann mit unserem webbasierten vicSDC-Tool durch einen Laien erfolgen.



Grundsätzlich können für jeden Befehl Beispielsätze in Form von Text vorgegeben werden. Je mehr Beispielsätze für einen Befehl angegeben werden, desto mächtiger und robuster gegen Variationen wird die Spracherkennung. Innerhalb eines Befehls können sogenannte Slots - Schlüsselwörter, Parameter und Wertebelegungen – durch manuelle Markierung definiert werden. Diese Slots werden der Sinnanalyse, der semantischen Interpretation des erkannten Satzes, zugrunde gelegt.

Die Sprachdialog-Gestaltung verknüpft die Beispielsätze mit den auszulösenden Befehlen und definiert Sprachausgaben des Systems. So kann der SUL mitgeteilt werden, auf welche Weise nach den einzelnen Bestandteilen (Slots und Values) von Befehlen (Intents) gefragt

werden soll, um den Befehl für eine Auslösung einer Maschinenfunktion zu vervollständigen (z.B. Auswahl eines Programms mit 3 Parametern), oder Mehrdeutigkeiten im Sprachbefehl aufzulösen.

Wakeup phrase: "Hello printing machine"		
Intent	Slot	Value
set output format	width	297
	height	420
	speed	1200
	instances	400

Weitere Funktionen einer Sprachausgabe können die kontextabhängige Bereitstellung akustischer Hilfefunktionen oder die akustische Meldung von Maschinenzuständen oder Systemfehlern sein.

Iterativ zur Validierung und Kontrolle des Designprozesses oder final zur Erzeugung eines die kundenspezifische SUL repräsentierenden Ressourcenpaketes, können im vicSDC Validierungs- und Trainingsschritte ausgelöst werden.

Wortwahl, Struktur des Satzes und Funktion einzelner Inhaltswörter werden in diesem Trainingsschritt zu einer Datenbank verarbeitet, welche die Sprachstruktur repräsentiert (Language Model), sowie zu einer Datenbank, die die Grundlage der semantischen Analyse bildet. Weitere Datenbanken entstehen automatisch für die Sprachausgaben sowie die Steuerschnittstelle der Zielanwendung. Im Ergebnis des Designprozesses wird durch das vicSDC-Tool ein Ressourcenpaket zur Verfügung gestellt, welches lokal auf dem Spectra Power-Box-PC zusammen mit der vorinstallierten, generischen Sprachdialog-Applikation von vicCONTROL die individuelle Steuerung repräsentiert.

## Können vorhandene Anwendungen mit Sprachinteraktion "nachgerüstet" werden?

Als Voraussetzung für eine nachträgliche Erweiterung durch eine Sprachbedienung sollte die zu steuernde Anwendung einen Zugriff auf Parameter, interne Zustände bzw. Funktionen erlauben. Bei hinreichender Abstraktion von Anwendung und Bedienung ist diese Voraussetzung meist gegeben.

Bei monolithisch implementierten Lösungen, bei denen Anwendung und User Interface eng miteinander verwoben sind, ist eine nachträgliche Erweiterung aufwändig, da solche Schnittstellen an geeigneter Stelle in die Anwendung eingebracht werden müssen.

Um eine kundenspezifische Sprachbedienlösung zu erstellen, wird zunächst der Prozessablauf analysiert und sämtliche Usecases zusammengetragen. Im Idealfall verfügt die Zielanwendung bereits über eine GUI, die bezüglich der Verwendung von Sprachinteraktion analysiert werden kann. Eine Prozessanalyse gibt Aufschluss über die akustischen Randbedingungen und ermöglicht uns, Empfehlungen für die Ausstattung mit Mikrofonen und Lautsprechern zu geben. Weiterhin erlaubt eine solche Prozessanalyse frühzeitig die Einbeziehung der Anwenderwünsche und die Erfassung des zu erwartenden Rationalisierungspotentials und somit - in Gegenüberstellung zur Investitionssumme - eine Bewertung der Anwendung unter Aufwand-Nutzen-Gesichtspunkten.

Eine frühzeitige Einbeziehung der Anwender beim Dialog-Design und bei der Prozessintegration sowie eine Diskussion der Vorteile mit dem Anwender, unterstützt beim Abbau von Vorurteilen und dem Überwinden von Akzeptanzhürden noch vor Einführung der Lösung.



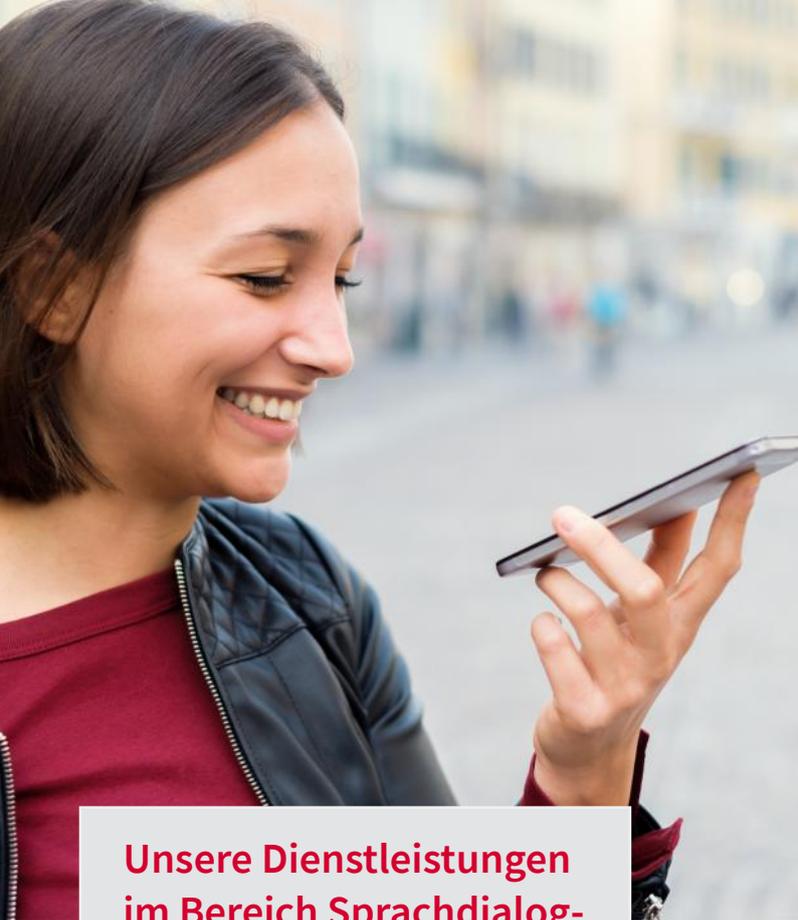
## Wo kann ich als Anwender Unterstützung erhalten?

Anwender, die sich für den Einsatz einer Sprachbedienung interessieren, wird empfohlen, zunächst mit der Spectra PowerBox 100-Voice von Spectra und dem vicSDC-Tool erste Erfahrungen zu sammeln und zunächst eine eigene Lösung nach dem Beispiel der „Tank-Demo“ zu entwickeln, die der Spectra PowerBox 100-Voice als Beispiel beiliegt.

Nutzer, die die Grundlagen des Designs einer kundenspezifischen Sprachbedienung erlernen und mit unseren Experten fachspezifische Fragen erörtern möchten, können an den nach Bedarf durch VIC durchgeführten Lehrgängen in Dresden teilnehmen. Diese Lehrgänge vermitteln Schritt für Schritt die Vorgehensweise beim Design eines Sprachdialogs und bei der Integration in die Zielanwendung.

Für Kunden, die Fragen zur Umsetzung der Sprachbedienung in ihrem Produkt haben, bietet VIC eintägige Fachworkshops an. Dort werden Aspekte der Bedienergonomie sowie akustische Randbedingungen diskutiert und gemeinsam die Systemarchitektur des Bedienkonzeptes anhand der Anforderungen des Kunden sowie die Umsetzungsschritte erarbeitet.

Kunden, die bereits konkrete Vorstellungen zur Umsetzung einer Sprachbedienung in ihrem Produkt haben, und die die Unterstützung unserer Fachexperten im Bereich Spracherkennung, Akustik, Hard- und Softwareentwicklung wünschen, bietet VIC die Durchführung von Entwicklungsprojekten an.



## Unsere Dienstleistungen im Bereich Sprachdialog- entwicklung

- Definition umfangreicher Sprachdialoge.
- Portierung der SUI auf eine (kunden-spezifische) Zielplattform und Integration mit der Zielanwendung.
- Erstellung und Optimierung von Aktivierungsworten.
- Unterstützung beim akustischen Geräte-design der Zielanwendung (Anordnung und Auswahl von Lautsprechern und Mikrofonen für das Zielgerät).

Die notwendige IT-Infrastruktur erhält der Anwender von Spectra. Je nach Umfang und Anforderung der Sprachbedienung kann durchaus auch ein anderer Industrie-PC zum Einsatz kommen. Speziell bei der Kombination von GUI (HMI) und SUI (Voice) macht der Einsatz eines Panel PC Systems Sinn. Spectra findet hier im Dialog mit dem Anwender die geeignete Lösung.

Da die Anbindung der Sprachsteuerung an die Automationsebene mittels MQTT Signal realisiert wird, ist eine stabile und sichere Netzwerkverbindung notwendig. Speziell hierfür bietet Spectra eine große Auswahl an Netzwerkprodukten für die Industrie, wie z.B. Ethernet Switchs und Mediakonverter.

Sollte die Maschine oder Anlage kein MQTT sprechen und verstehen können, dann helfen MQTT-fähige Ein- und Ausgabe-Module von Spectra weiter. Diese industriellen Schaltschrankmodule werden einfach an das Ethernet angeschlossen und kommunizieren auf der einen Seite per MQTT-Kommandos und auf der anderen Seite mit elektrischen Signalen, wie digitale und analoge Ein- und Ausgänge.

Und treu nach unserem Motto „powered by individuality“ ist ein wesentlicher Schwerpunkt die Realisierung von kundenspezifischen Projekten. Hierbei spielt die kompetente Beratung eine wichtige Rolle, die wir mit unserem technischen Vertrieb vor Ort im ganzen DACH-Gebiet realisieren.



# KONTAKT

**voice INTER connect GmbH**



**voice INTER connect**

Ammonstraße 35  
01067 Dresden  
Deutschland

Ansprechpartnerin:  
Dr.-Ing. Diane Hirschfeld

Kontakt:  
Telefon: +49 351 407 526 50  
Telefax: +49 351 407 526 55  
E-Mail: [info@voiceinterconnect.de](mailto:info@voiceinterconnect.de)



Mahdenstraße 3  
72768 Reutlingen  
Deutschland

Ansprechpartner:  
Klaus Rottmayr

Kontakt:  
Telefon: +49 7121 1432 10  
Telefax: +49 7121 1432 190  
E-Mail: [spectra@spectra.de](mailto:spectra@spectra.de)

# KONTAKT

**Spectra GmbH & Co. KG**

## Bildnachweise

<b>Seite</b>	<b>Quelle</b>
Titel	© phonlamaipphoto – stock.adobe.com
1	© phonlamaipphoto – stock.adobe.com
2	© Tyler Olson – stock.adobe.com
3	PowerBox 100-Voice / Spectra GmbH & Co. KG
4	vicDIVA / voice INTER connect GmbH © Kzenon – stock.adobe.com
5	© photostriker – stock.adobe.com
6	© WavebreakMediaMicro – stock.adobe.com
7	voice INTER connect GmbH / Spectra GmbH & Co. KG
8	© Wire_man – stock.adobe.com
9	© REDPIXEL – stock.adobe.com
10	© Paolese – stock.adobe.com © momius – stock.adobe.com