

# Sprachsteuerung für Mess- und Prüfgeräte

Constanze TSCHÖPE<sup>\*</sup>, Dieter JONEIT<sup>\*</sup>, Frank DUCKHORN<sup>\*\*</sup>, Rüdiger HOFFMANN<sup>\*\*</sup>,  
Guntram STRECHA<sup>\*\*</sup>, Matthias WOLFF<sup>\*\*</sup>

<sup>\*</sup> Fraunhofer IZFP Dresden, Maria-Reiche-Straße 2, 01109 Dresden  
constanze.tschoepe@izfp-d.fraunhofer.de

<sup>\*\*</sup> TU Dresden, Institut für Akustik und Sprachkommunikation

**Kurzfassung.** Überall, wo Mess- und Prüfgeräte eingesetzt werden, beispielsweise für die zerstörungsfreie Prüfung mittels Ultraschall bei der Flugzeuginspektion, sind Einstellungen vorzunehmen. Es gibt typische Anwendungsfälle, in denen der Prüfer nicht in der Lage ist, das Gerät selbst manuell zu bedienen (Enge, Schmutz, prozessbedingte Umstände, wie z. B. eine Hand am Gerüst und eine Hand am Prüfkopf, oder die Notwendigkeit des Tragens von Handschuhen). Dann kann eine Sprachsteuerung Abhilfe schaffen. Das erfordert einerseits, dass der Prüfer alle Einstellungen per Sprache vornimmt (Spracherkennung), und andererseits, dass das Gerät in der Lage ist, die vorgenommenen Befehle zu quittieren und eine Rückmeldung zu geben (Sprachsynthese).

Dafür wurde ein Sprachsteuerungssystem entwickelt, das die Eingabe "fließender Sprache" (d. h. möglichst natürlichsprachlicher Äußerungen wie "Messung starten in 5 Sekunden für 1,5 Minuten") und dadurch eine Einstellung der notwendigen Parameter erlaubt. Rechenintensive Algorithmen wurden aus Gründen der Energieeffizienz in programmierbare Logikschaltkreise ausgelagert. Außerdem gelang es, einen neuartigen integrierten Ansatz zur Spracherkennung und -synthese umzusetzen. Das Gerät wurde unter Feldbedingungen und in unterschiedlichsten Umgebungen (in geräuschintensiven Werkhallen, an Straßenkreuzungen etc.) erprobt. Die zugrunde liegende reguläre Grammatik kann an Kundenbedürfnisse und damit an Mess- und Prüfgeräte für jegliche Anwendung angepasst werden.

## 1. Einführung

Beim Einsatz von Mess- und Prüfgeräten, beispielsweise für die zerstörungsfreie Prüfung mittels Ultraschall bei der Flugzeuginspektion, müssen Einstellungen vorgenommen werden. Ist der Prüfer nicht in der Lage, das Gerät selbst manuell zu bedienen (Enge, Schmutz, prozessbedingte Umstände, wie z. B. eine Hand am Gerüst und eine Hand am Prüfkopf, oder die Notwendigkeit des Tragens von Handschuhen), kann eine Sprachsteuerung Abhilfe schaffen. Das erfordert einerseits, dass der Prüfer alle Einstellungen per Sprache vornimmt (Spracherkennung), und andererseits, dass das Gerät in der Lage ist, die vorgenommenen Befehle zu quittieren und eine Rückmeldung zu geben (Sprachsynthese).

## 2. Ziele

Ziel des Projektes war es, eine Sprachsteuerung für Mess- und Prüfgeräte (speziell Ultraschallprüfgeräte) zu entwickeln. Das bedeutet, dass alle Einstellungen am Gerät, die bisher nur mit Hilfe einer Bedienoberfläche erfolgten, auch durch Sprache vorgenommen werden können. Dabei sollte möglichst natürliche Sprache ein- und ausgegeben werden.



Die Sprache unterliegt einer regulären Grammatik und muss bestimmte Regeln einhalten. Eine typische Sprachäußerung ist beispielsweise „Messung in 10 Sekunden für 0,5 Sekunden“. Das Hardwaremodul sollte über die USB-Schnittstelle an das Prüfgerät anschließbar und mittels Headset steuerbar sein. Um die Ausrüstung für den Prüfer im Außenfeld möglichst gering und handhabbar zu halten, sollte die Hardware miniaturisiert und im Sinne einer optimalen Arbeit der Stromverbrauch minimiert werden.

### **3. Technologie**

#### *3.1 Spracherkenner und -synthesator*

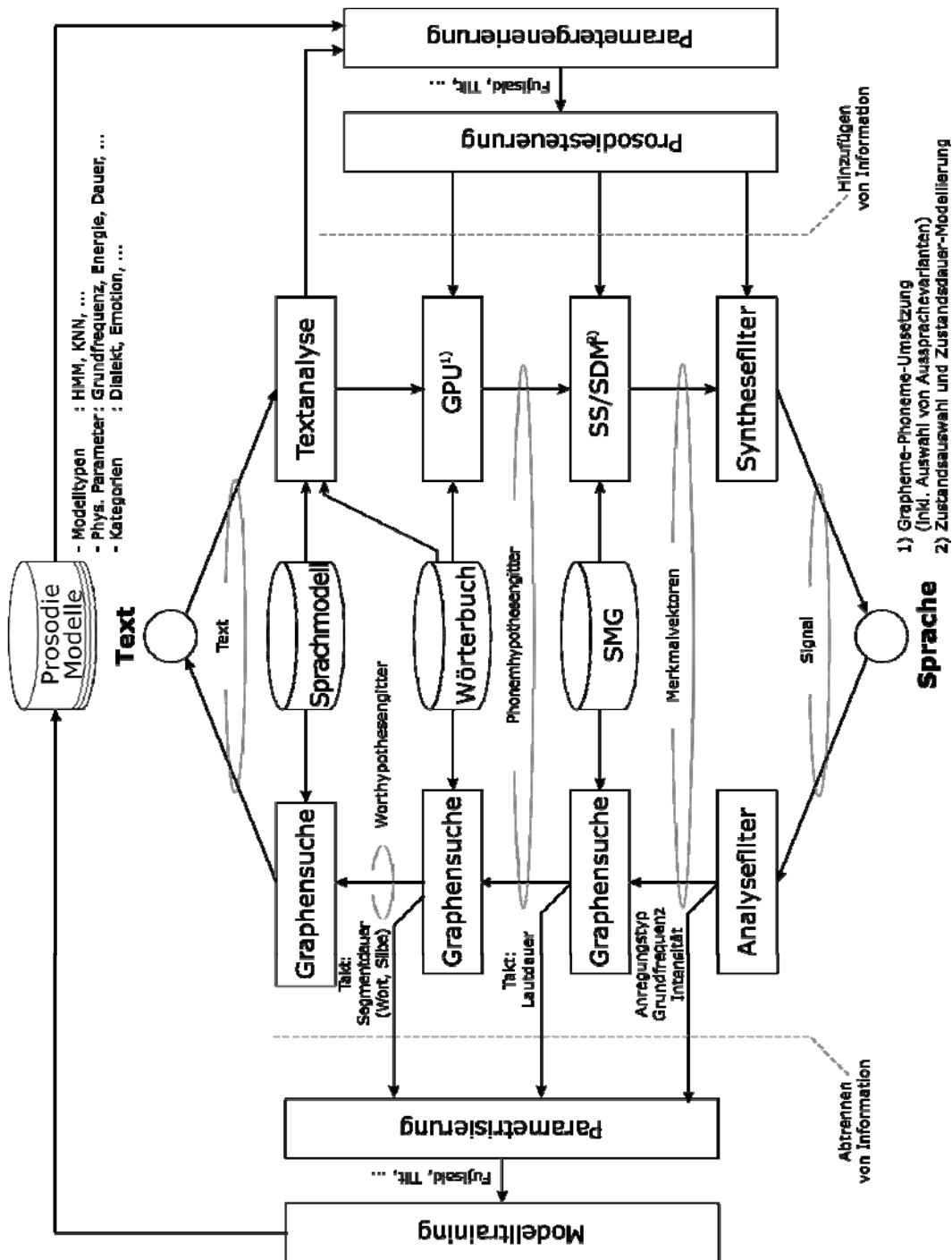
Als Technologie wurde das am Institut für Akustik und Sprachkommunikation entwickelte UASR (Unified Approach to Speech Synthesis and Recognition) verwendet. Das Blockschaltbild (Abbildung 1) zeigt die Funktionsweise. Der linke Pfad stellt die Spracherkennung, der rechte die Sprachsynthese dar.

Bei der Spracherkennung wird zuerst die aufgenommene Sprache analysiert, und sprachliche Eigenschaften (wie Satzmelodie, Emotion) – Prosodie genannt – werden abgetrennt. Jede Ebene wird durch Modelle (Wörterbuch, Sprachmodell etc.) unterstützt. Verschiedene hierarchische Ebenen verdeutlichen die Erhöhung der Abstraktion von unten nach oben. Beim Durchlaufen der Ebenen wird aus dem Sprachsignal letztendlich der Text gefiltert. Mit Hilfe des Wörterbuchs entscheidet das System, ob das Wort enthalten und damit gültig ist. Das Sprachmodell überprüft die Gültigkeit der getroffenen Aussage (aufeinanderfolgende Wörter).

Die Sprachsynthese kann als Umkehrung der Spracherkennung angesehen werden. Sie durchläuft die hierarchischen Ebenen von oben nach unten, greift auf die gleichen Modelle wie die Spracherkennung zu und erzeugt aus dem Text eine sprachliche Aussage. Um der Sprache einen natürlichen Klang zu verleihen, wird die Prosodie wieder hinzugefügt.

#### *3.2 Fließende Sprache*

Um das System zu aktivieren, ist ein Codewort erforderlich. Bei der Auswahl des Codewortes wurde darauf geachtet, dass es im natürlichen Sprachgebrauch selten vorkommt. Befehle zum Starten und Stoppen einer Messung in verschiedenen Varianten und zum Steuern des Prüfgerätes müssen in der Grammatik enthalten sein. Damit ist der Prüfer nicht auf eine bestimmte Wortwahl festgelegt, die er auswendig kennen müsste, sondern er kann ein- und denselben Befehl auf verschiedene Weise ausdrücken, was eine wesentliche Voraussetzung für die Akzeptanz des Gerätes darstellt. Damit das Sprachsteuerungssystem durch die Prüfer angenommen wird, muss es in der Lage sein, die „Sprache der Prüfer“ zu verstehen. Die Festlegung aller möglichen Wendungen (Wortschatz) erfolgte daher in Zusammenarbeit mit den Prüfern. Um dem Anwender eine Rückmeldung zu geben, wurden 3 verschiedene Earcons (akustische Elemente) verwendet: positive Befehlsquittung (Befehl wurde erfolgreich erkannt), negative Befehlsquittung (Befehl wurde nicht erkannt) und Sleep (System wurde in den Schlafmodus versetzt). Um alle möglichen Zahlen inkl. Nachkommastellen einstellen zu können, wurde eine Zahlwortgrammatik implementiert.



**Abbildung 1:** Blockschaltbild des UASR [1]. Der linke Pfad verdeutlicht die Spracherkennung, der rechte Pfad die Sprachsynthese. Die einzelnen hierarchischen Ebenen greifen dabei auf die gleichen Modelle zu.

### 3.3 Umsetzung

Ein neuartiger Ansatz des Projekts bestand in der Umsetzung der Sprachsteuerungsalgorithmen in programmierbare Logikschaltkreise. Die Algorithmen zur Spracherkennung konnten erstmals fast vollständig in Hardware verlegt werden, was enorme Vorteile bezüglich der Rechenleistung und der Energieeffizienz bietet. Gemeinsam mit den Projektpartnern konnte ein Demonstrator gefertigt werden, bei dem alle Bauelemente auf einer Platine vereinigt sind.




#### 4. Feldtests

Bei der Sinus Messtechnik GmbH Leipzig wurden umfangreiche Erprobungen unter Feldbedingungen durchgeführt. Insgesamt wurden in einem ersten Feldtest ca. 5 Stunden Audiodaten, darunter 42 Minuten Sprache, aufgezeichnet. 6 verschiedene Sprecher (5 männliche und 1 weiblicher) führten die Aufnahmen durch. In folgenden Umgebungen wurden die Tests durchgeführt:







- Büro
- im Auto (gute und schlechte Straßen gemischt)
- stark befahrene Kreuzung
- Werkstatt mit CNC-Fräse (85-95 dB Störung)
- Nebenraum der Werkstatt (60-65 dB Störung)
- Garage neben Werkstatt.

Insgesamt waren 9 Millionen verschiedene Eingaben möglich. Die Ergebnisse sind in Abbildung 2 dargestellt. Der grüne Balken verdeutlicht eine erfolgreiche Erkennung. Die aufgetretenen Fehler wurden untersucht und verschiedenen Klassen zugeordnet. Grob können sie in System- (schmale, rote Balken in Abbildung 2) und Benutzerfehler (breitere, farbige Balken) unterteilt werden:

##### Systemfehler

	REC	Erkennungsfehler
	REJ	Rückweisungsfehler
	VAD	Sprach-/Pausedetektionsfehler

##### Benutzerfehler

	GAIN	Signalpegel unter Schwellwert
	SHORT	Signallänge unter Schwellwert
	JOIN	Mehrere Befehle zu schnell hintereinander
	SEP	Zu große Pause innerhalb eines Befehls
	CMDERR	Kein gültiger Befehl
	SLEEP	Befehl im Schlafmodus

	Erfolgreiche Erkennungen
---	--------------------------

Die besten Ergebnisse liefert natürlich die ruhige Büroumgebung. Abgesehen von der extrem lauten Werkstatt sind die Ergebnisse gut.

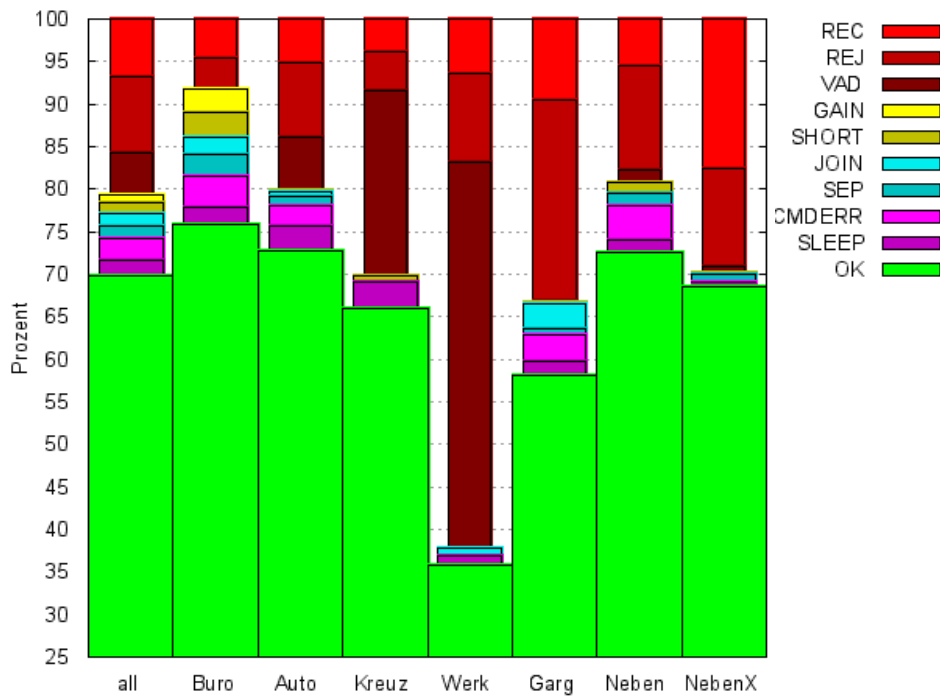


Abbildung 2: Ergebnisse des Feldtests I.

Um die Ergebnisse weiter zu verbessern, wurde eine Störgeräuschunterdrückung implementiert. Sie dient dazu, Nebengeräusche aus dem Signal zu filtern. Anschließend erfolgte ein zweiter Feldtest, bei dem ca. 1 Stunde Audiodaten aufgezeichnet wurden. Dabei standen 3 Sprecher zur Verfügung. Die Aufnahmen wurden in zwei Umgebungen durchgeführt:

- im Auto
- Nebenraum der Werkstatt.

Abbildung 3 zeigt die Ergebnisse. Besonders im Nebenraum der Werkstatt konnte eine deutliche Verbesserung erzielt werden. Die Erkennungsraten stiegen von 73 % auf 81 %.

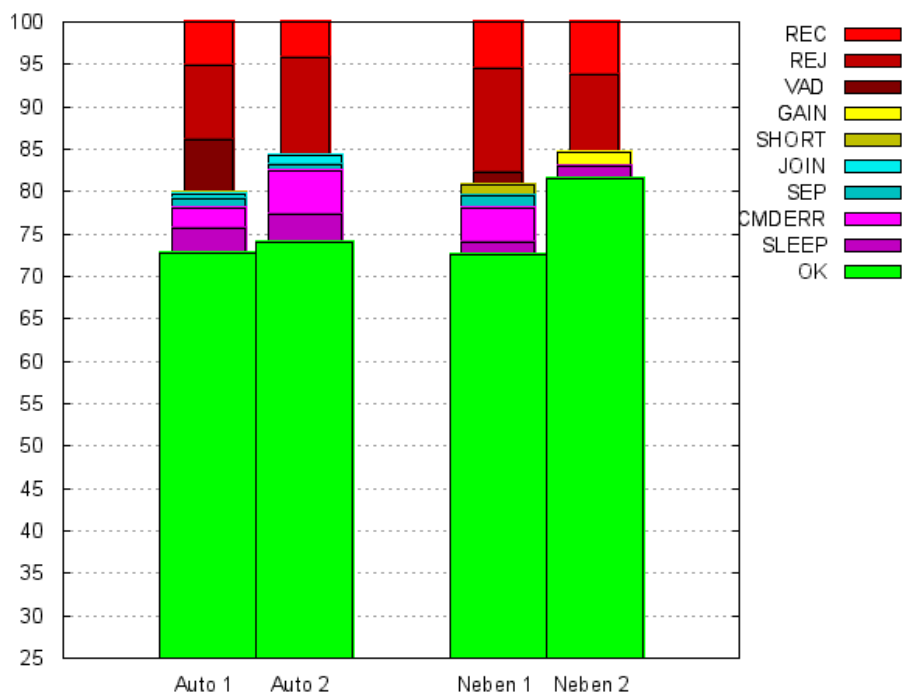


Abbildung 3: Ergebnisse des Feldtests II.

## 5. Ausblick

Das vorgestellte Sprachsteuerungssystem ist nicht nur für Mess- und Prüfgeräte geeignet. Es kann für Geräte jeder Art, die manuell einstellbar und über eine USB-Schnittstelle fernsteuerbar sind, angepasst werden. Die Grammatik kann auf spezielle Kundenwünsche zugeschnitten werden. Außerdem ist das entstandene System nicht nur auf Sprachsignale ausgelegt. Auch Anwendungen aus dem Gebiet der zerstörungsfreien Prüfung, wie die Qualitätsbewertung von Bauelementen, die Zustandsüberwachung von Eisenbahnrädern oder Flugzeugbauteilen, sind mit diesem System lösbar. Die Sprachmodelle werden dann durch Datenmodelle, die in einem Anlernprozess erzeugt werden, ersetzt [10].

## Dank

Unser Dank gilt dem VDI Technologiezentrum Düsseldorf und dem Bundesministerium für Bildung und Forschung, da sie uns dieses Projekt durch ihre Zuwendung ermöglicht haben.

## Referenzen

- [1] M. Eichner, M. Wolff, R. Hoffmann: A unified approach for speech synthesis and speech recognition using Stochastic Markov Graphs. In Proceedings of the International Conference on Spoken Language Processing 2000. ICSLP 2000, volume 1, pages 701–704, Beijing, PR China, Oct. 2000.
- [2] C. Tschöpe, M. Wolff, D. Joneit: Sprachsteuerung von Mess- und Prüfgeräten. Tagungsband zur BMBF-Statustagung „KMU-innovativ: IKT“, 1 Seite, November 2008.
- [3] F. Duckhorn, M. Wolff, G. Strecha, R. Hoffmann: An application example for unified speech synthesis and recognition using Hidden Markov Models. One day meeting on unified models for speech recognition and synthesis, Mar. 2009, Birmingham, U.K, 2009.
- [4] C. Tschöpe, M. Wolff: SSMG-Sprachsteuerung für Mess- und Prüfgeräte. Im Rahmen der Ausstellung des Materialforschungsverbundes Dresden zur Langen Nacht der Wissenschaften, 19.6.2009.
- [5] G. Strecha, M. Wolf, F. Duckhorn, S. Wittenberg, C. Tschöpe: The HMM Synthesis Algorithm of an Embedded Unified Speech Recognizer and Synthesizer. In Proc. Interspeech 2009, 10th Annual Conference of the International Speech Communication Association, 1763-1766, Brighton 2009, 6.-10.9.2009.
- [6] F. Duckhorn, G. Strecha, M. Wolff, R. Hoffmann. Ein Sprachdialogsystem mit begrenzten Hardwareressourcen. In R. Hoffmann, editor, Elektronische Sprachsignalverarbeitung 2009. ESSV 2009, Volume 52 of Studentexte zur Sprachkommunikation, Seite 883, Dresden, Germany, Sep. 2009. TUDpress, Dresden. ISBN: 978-941298-31-6.
- [7] M. Wolff, C. Tschöpe: Pattern Recognition for Sensor Signals. In Proc. IEEE SENSORS 2009 Conference, 665-668, Christchurch, Neuseeland, 25.-28.10.2009.
- [8] C. Tschöpe, D. Joneit, M. Wolff, G. Strecha, F. Duckhorn, R. Hoffmann: Sprachsteuerung für Mess- und Prüfgeräte. Innovationsforum „Software Saxony“, Dresden, 23.04.2010.
- [9] F. Duckhorn, G. Strecha, M. Wolff, R. Hoffmann: Entwicklung und Performance eines Sprachdialogsystems mit begrenzten Hardwareressourcen, ESSV 2010, Berlin, 6.-8. Sep. 2010.
- [10] C. Tschöpe, M. Wolff: Statistical classifiers for structural health monitoring. IEEE Sensors Journal, 9(11):1567–1676, Nov. 2009.