



Anmerkungen zur Effizienz der Qualitätssicherung in Sprachapplikationen

Nürnberg, 06.10.2009 - Kerstin Sehnert - atip GmbH

Wirtschaftliche Kriterien

- kurze Amortisierungsdauer
- Investitionssicherheit
- effizienter Betrieb

Technische Kriterien

- fehlerfreie Funktion
- hohe Verfügbarkeit
- geringe Latenzen
- Skalierbarkeit

Usability-Kriterien

- attraktives Funktionsangebot
- intuitive Menüstruktur
- verständliches Prompting
- postive Kommunikationserfahrung



Leitfaden „Qualitätskriterien, Maß und Verfahren für Sprachapplikationen“

Initiative Voice Business / Fraunhofer IAO / voicecommunity.de (2007-2009)

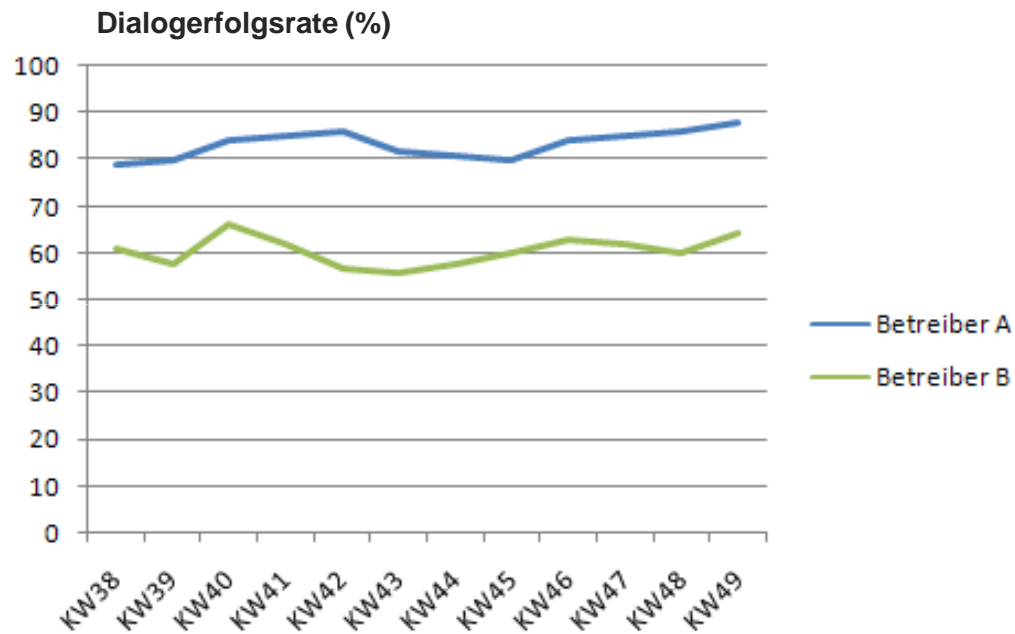
Usability-Testing: Verfahren

- 1) Auswahl einer repräsentativen Gruppe potentieller Benutzer
- 2) Definition von Testcases bzw. User Tasks
- 3) Durchführen der Tests
- 4) Befragung der Tester
- 5) Analyse von Logfiles, Anrufmitschnitten, Utterances
- 6) Benennung und Durchführung von Optimierungsmaßnahmen

Usability-Testing: Probleme

- Tester folgen **Instruktionen**
- **Situation und Motivation** der Benutzung lassen sich nicht simulieren

Beispiel aus der Praxis

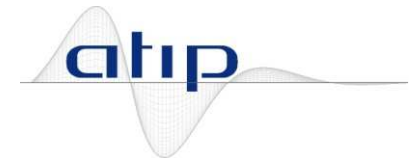


- nahezu identische Applikationen
- ähnliche Nutzergruppen
- gleiche Rahmenbedingungen

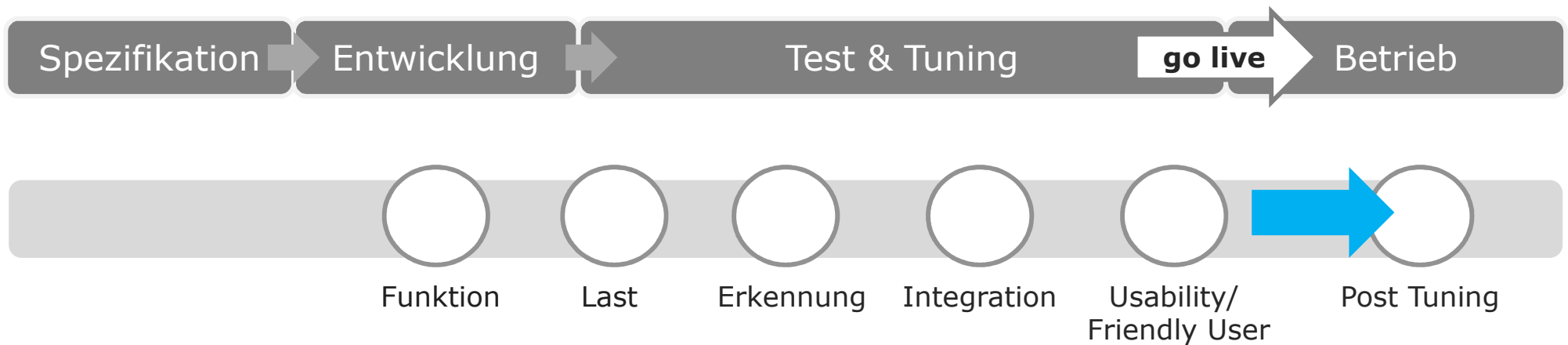


Präsent als Motivation
(Betreiber A)

Usability im Test: Probleme



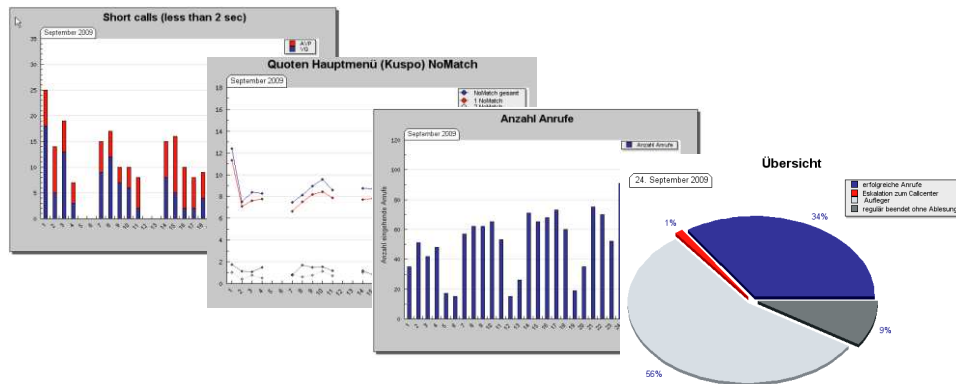
- Tester folgen **Instruktionen**
- **Situation und Motivation** der Benutzung lassen sich nicht simulieren



Verlagerung in das Post Tuning

- intensives Monitoring im Roll Out-Prozess

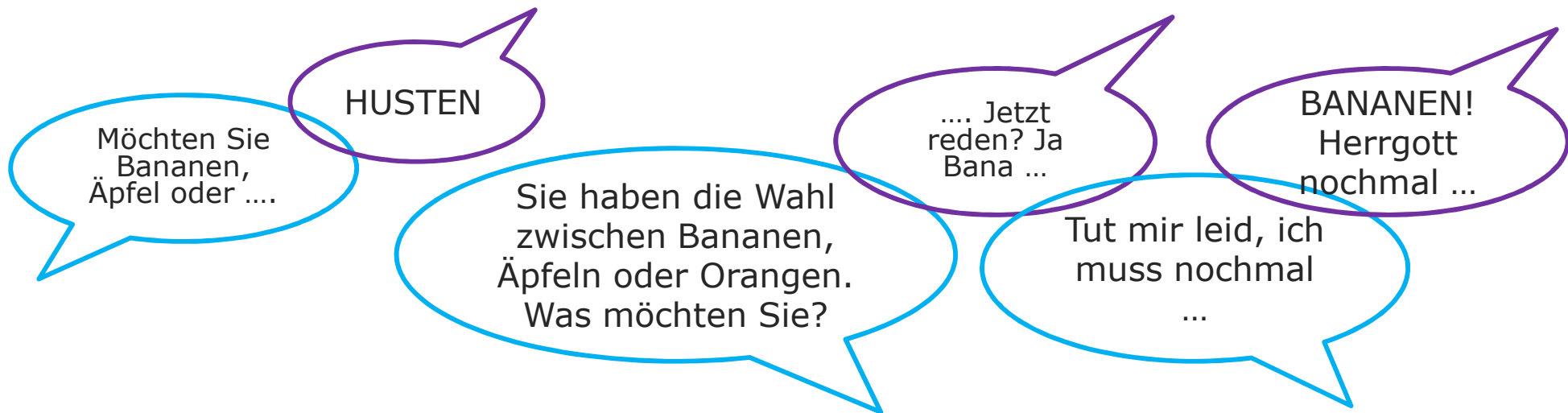
keep track! voice application monitoring and reporting framework



- Entwicklung von Anrufzahlen und Erfolgsraten
- wo brechen Anrufer ab?
- an welcher Frage häufen sich Fehler?

- Analyse korrespondierender Daten
- Ableiten und Durchführen von Optimierungsmaßnahmen

Praxisbeispiel: Barge-In-Steuerung & Eingabesignalisierung



- Barge-In ja/nein?
- Robustheit versus Komfort
- Designrichtlinien: kein Wechsel des Eingabemodus

Bitte nennen Sie ihre 6-stellige
Kundennummer <beep>

Ihre Kundennummer lautet also
457238, ist das richtig? <beep>

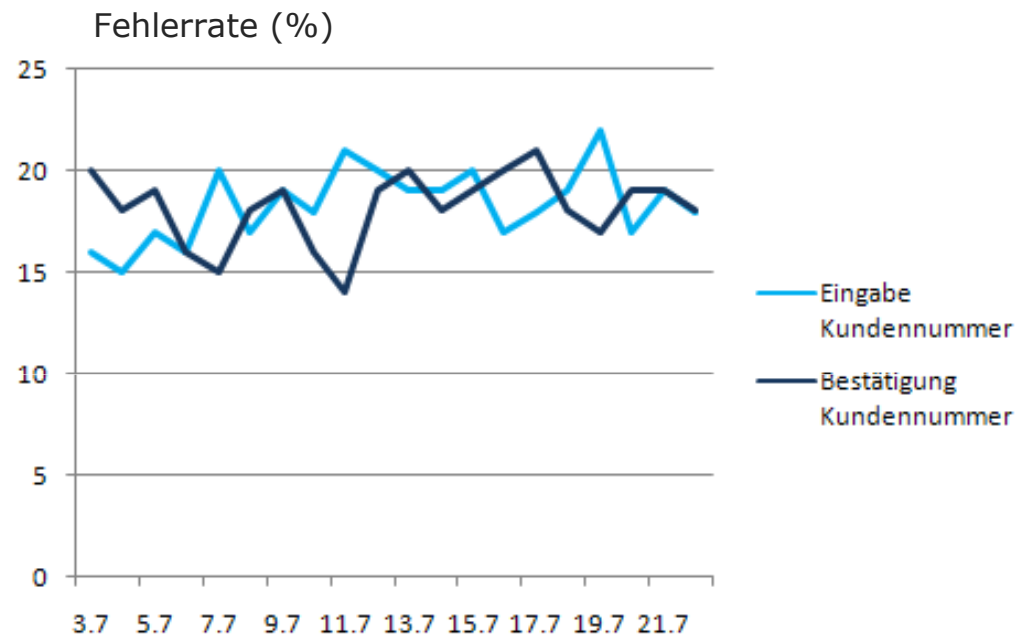
457238

ja

- Ziffernketteneingabe versus Ja-/Nein-Frage
- höhere Fehlerrate wurde bei der Nummerneingabe erwartet
- Friendly User Tests bestätigten das

Bitte nennen Sie ihre 6-stellige
Kundennummer <beep>

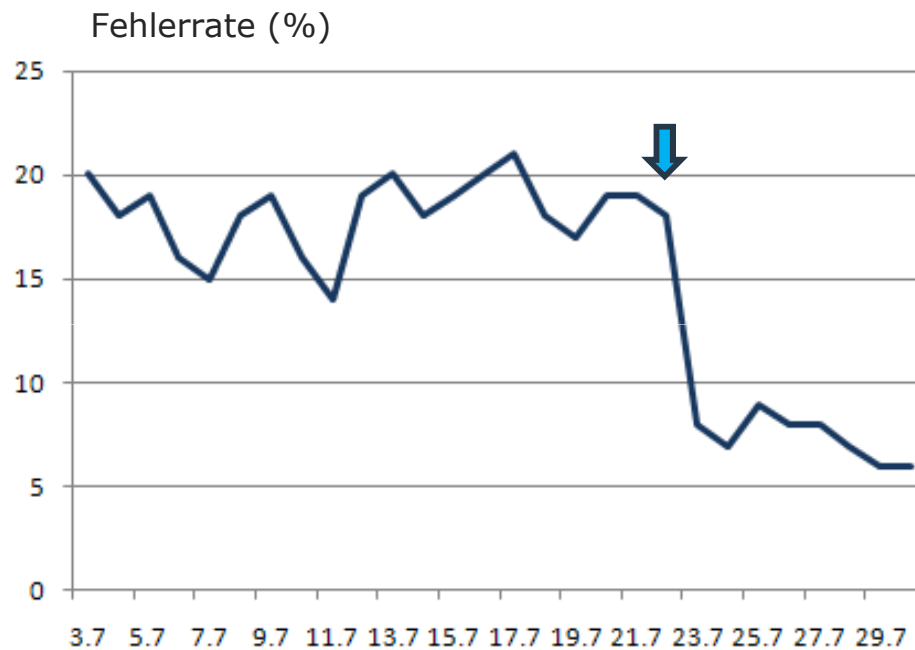
Ihre Kundennummer lautet also
457238, ist das richtig? <beep>



- Livebetrieb: unerwartet hohe Fehlerraten bei der Ja-/Nein-Frage

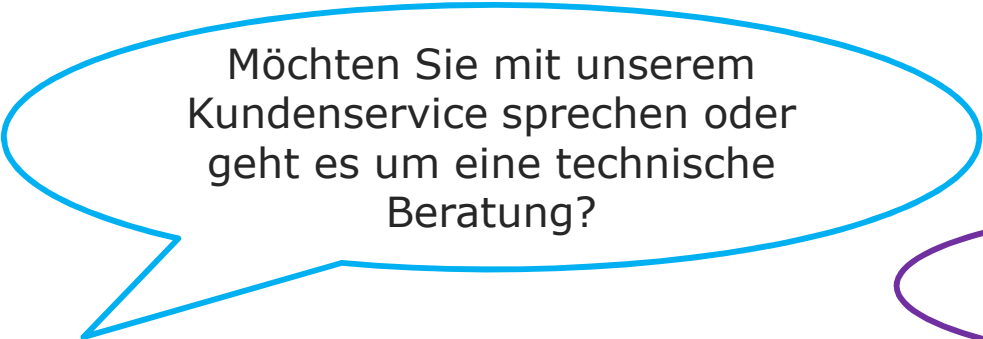
- Unterbrechungsversuche führten zu einer Häufung von nomatch und noinput
- 2000 Anrufe im Friendly User Test
- Tester verhielten sich diszipliniert
- kein Hinweis auf Probleme

Ihre Kundennummer lautet also 457238, ist das richtig? <beep>

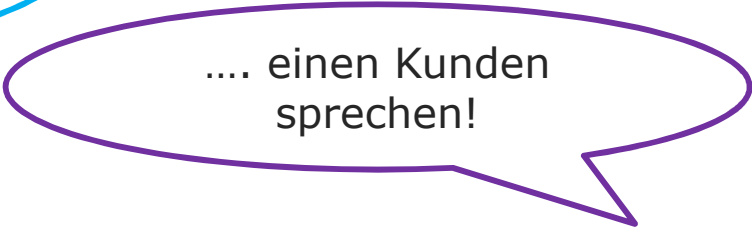


- Wechsel des Eingabemodus
- Sinken der Fehlerrate

Praxisbeispiel: Konfrontation mit unerwarteten Benutzeräußerungen

A blue speech bubble with a tail pointing towards the bottom left. It contains the text: 'Möchten Sie mit unserem Kundenservice sprechen oder geht es um eine technische Beratung?'

Möchten Sie mit unserem Kundenservice sprechen oder geht es um eine technische Beratung?

A purple speech bubble with a tail pointing towards the bottom right. It contains the text: '.... einen Kunden sprechen!'

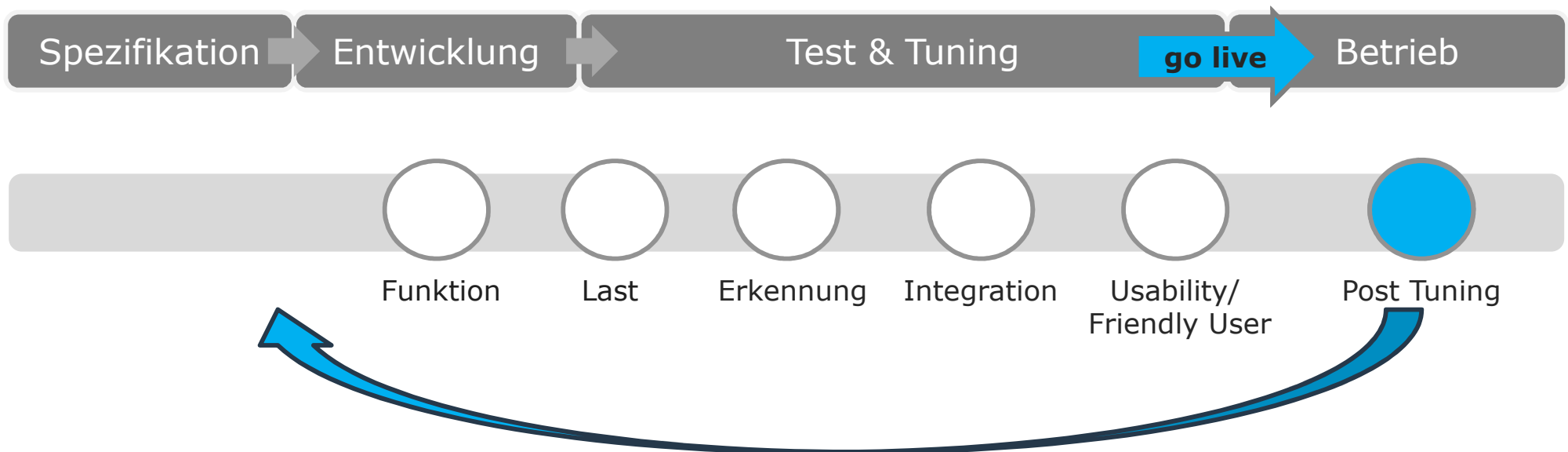
.... einen Kunden sprechen!

- keine Variante dieser Formulierung wurde in Tests genutzt
- eine Optimierung auf Basis von Livedaten senkte die Fehlerrate um 60%

Diskrepanz zwischen Erkenntnissen aus Tests und Post Tuning

Post Tuning: unverzichtbares Mittel auf dem Weg zur optimalen Applikation

Effizienz: Erreichen größtmöglicher Qualität mit möglichst kleinem Aufwand



Probleme:

- Nutzer wird als Tester missbraucht
- Betreiber launchen noch nicht optimale Applikationen
- Änderungen im Produktivbetrieb sind aufwendig

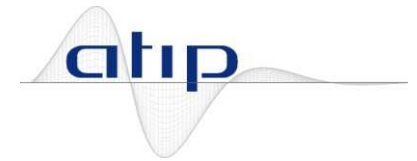
... Applikationen sind schon bei Go Live optimal

... es gibt messbare Anhaltspunkte, ob Designkriterien, die für Anwendung X richtig sind, in Anwendung Y ebenso gelten

... es ist weniger aufwendig, Usability-Kriterien zu testen



Effizienzsteigerung durch Automatisierung



Lösungsansatz: Effizienzsteigerung durch Automatisierung

- ohne menschliche Tester
- messbar und vergleichbar
- schnell und reproduzierbar
- in großen Mengen



Automatisierte Bewertung von Usability-Aspekten? Der VUI wundert sich ...



Workbench für automatisierte Usability-Tests

- Idee: simuliert Benutzerverhalten unter Verwendung mentaler Modelle
- Definition prototypischer Nutzergruppen
- Ableitung typischer Nutzungsverhalten und -parameter
- Simulation der Interaktionen zwischen Benutzer und System
- Prognosen zur Gebrauchstauglichkeit



Workbench für automatisierte Usability-Tests

Berücksichtigung von Usability-Kriterien bestimmter Nutzergruppen
MeMo kann helfen ...

- problematische Dialogpfade zu finden
- verschiedene Designoptionen zu vergleichen
- Fehlerbehandlungsstrategien zu bewerten

Wir stehen am Anfang eines Evaluationsprojektes

Zusammenfassung

Stand der Dinge

Aktuell ist Post-Tuning ein unverzichtbares Mittel für die Optimierung von Usability-Kriterien

Damit sind wir nicht zufrieden ...

Optimierung im Betrieb und Tests am echten Benutzer sind nicht effizient

Wohin geht der Weg?

Ansatz zur Effizienzsteigerung durch frühzeitig automatisierte Tests
Ist MeMo für den Praxiseinsatz geeignet?

Vielen Dank



atip GmbH
Daimlerstraße 32
D-60314 Frankfurt
Tel +49 69 710407 0
Fax +49 69 710407 124
www.atip.de