EIN INTERAKTIVES TUTORIAL FÜR EIN SPRACHBEDIENSYSTEM IM FAHRZEUG

Franziska Höhne, Bärbel Jeschke*, Klaus Fellbaum***

**Harman/Becker Automotive Systems, **BTU Cottbus*
fhoehne@harmanbecker.com

Abstract: Aufgrund der steigenden Zahl zusätzlicher Funktionen im Fahrzeug bietet sich die Nutzung von Sprache als zusätzlicher Informationskanal für die Bedienung an. Bestehende Sprachbediensysteme litten bisher unter mangelnder Nutzerakzeptanz. Ein Ansatz, diese Situation zu verbessern, könnten geeignete Hilfen für den Nutzer sein. Ausgehend von einer Analyse wesentlicher Merkmale bereits existierender Hilfen wurde ein interaktives Tutorial entwickelt, das in seiner aktuellen Version in das *VoiceCar*-System von *Harman/Becker* integriert ist. Das Tutorial vermittelt dem Erstnutzer eines Sprachbediensystems in einer Einführung die Grundlagen der Bedienung. In ausgewählten Bedienbeispielen für Telefon, Navigation und MP3-Player kann sich der Anwender das Prinzip der Sprachbedienung ansehen und selbst ausprobieren. Dazu sind die Bedienbeispiele jeweils in die Lerneinheiten *Demo*, *Übung* und *Tipps* untergliedert. Zusätzliche Hinweise zu den Bedienbeispielen und der Sprachbedienung helfen dem fortgeschrittenen Anwender, das System effektiv zu nutzen. Zur Ersterprobung wurden Akzeptanztests in einer Testumgebung durchgeführt. In den Ergebnissen hat sich gezeigt, dass die Probanden das Tutorial als Lernmaterial annehmen. Die übersichtliche Menüstruktur, die von einer Tutorial-Stimme begleitet wird, und die eindeutigen Bezeichnungen der Lerneinheiten halfen den Probanden, sich im Tutorial zu orientieren und Informationen schnell wieder zu finden. Besonders zweckmäßig wurde auch die Gliederung in *passive* und *interaktive* Lerneinheiten sowie die Verwendung zwei weiterer Stimmen, der *System-Stimme* und der *Beispielnutzer-Stimme*, bewertet, wodurch die Bedienprinzipien der Sprachbedienung besonders klar präsentiert werden können.

1 Einführung und Aspekte eines interaktiven Tutorials

Warum soll für ein Sprachbediensystem (SBS) im Fahrzeug ein geeignetes Lernmaterial entwickelt werden? Schließlich besteht doch der Anspruch, dass ein solches System intuitiv nutzbar sein soll. Bekanntermaßen ist ein solches SBS im Fahrzeug jedoch in seiner Hardware beschränkt. Durch den begrenzten Raum kann nur eine begrenzte Speicherkapazität und Rechenleistung zur Verfügung gestellt werden. Beides hat zur Folge, dass Sprachbediensysteme im Fahrzeug keine natürlichsprachlichen Dialoge anbieten können, sondern auf kommandobasierte Dialoge zurückgreifen müssen [1]. Für den Nutzer eines solchen Systems bedeutet dies wiederum, dass er zur optimalen Nutzung bestimmte Kommandos beherrschen muss. Die gültigen Kommandos findet der Anwender im Handbuch zu seinem Fahrzeug. Doch wer liest sich freiwillig das Handbuch seines Fahrzeugs durch? Und wer lernt in seiner Freizeit gern Vokabeln? Zu bedenken ist auch, dass *Harman/Becker* überwiegend Sprachbediensysteme für Oberklassewagen herstellt. Die Besitzer und damit die Zielgruppe der Sprachdialogsysteme sind überwiegend ältere Personen, die neue Inhalte nicht mehr so schnell lernen wie ein junger Mensch. Es wird also eine Lernhilfe zu einem Sprachbediensystem gesucht, die dem Lernenden unterhaltsam und spielerisch die gültigen Kommandos vermittelt, ohne dass ihm bewusst wird, dass er gerade Vokabeln lernt.

Umfragen bei Autohändlern im Raum Ulm ergaben, dass das Bedienprinzip eine weitere Barriere darstellt. Der Autohändler selbst hat im Verkaufsgespräch keine Zeit, dem Kunden

das Sprachbediensystem zu zeigen. Ein Verkaufsgespräch stellt für den Kunden auch eine Stresssituation dar. Er muss sich viele Informationen merken, die wichtig für die sichere Fahrzeugführung sind. Die Funktionsweise eines Sprachdialogsystems ist in dieser Situation aus Sicht des Kunden und Autohändlers nebensächlich [2].

Als Anwender eines SBS muss man wissen, wie ein solches aktiviert wird, wann es einem zuhört und wann man sprechen kann. Ein geeignetes Lernmaterial sollte hier dem Anwender den Einstieg erleichtern, indem es das Bedienprinzip kombiniert auditiv und visuell präsentiert. Die Präsentation sollte deswegen multimodal sein, da das SBS mittels akustischer und visueller Zeichen dem Nutzer deutlich macht, wann er sprechen darf.

Bei der Nutzung eines SBS wird schnell die mit der Zahl der zu bedienenden Geräte wachsende Komplexität solcher Systeme deutlich, die den Anwender schnell überfordern kann. Ein Lernmaterial sollte dem Anwender nicht nur eine Übersicht über alle sprachbedienbaren Geräte geben. Es sollte dem Lernenden helfen, sich in die Menüstrukturen des Systems einzudenken. Er soll dabei lernen, dass das System ihn durch den Dialog führt, er also nur das sprechen muss, was das System von ihm verlangt.

Es gibt bereits Hilfen für Sprachbediensysteme, die das Bedienen eines SBS erleichtern sollen. Dazu gehören die integrierte Hilfefunktion sowie Audio-CDs und Lehrfilme wie für die Linguatronic von Mercedes-Benz. Die integrierte Hilfefunktion nutzt allerdings nur dem bereits versierten Anwender, der das Bedienprinzip und auch die Hilfefunktion schon kennt. Die Audio-CDs bieten umfangreiche Informationen, die aber nicht parallel ausprobiert werden können. Der Anwender muss sich die Abläufe merken, bis er die Möglichkeit hat, sie am Gerät auszuprobieren. Hier stößt man auf das Memorierungsproblem. Man kann sich nur 5 ± 2 Chunks merken [3][4], was bezogen auf die Informationsmenge, die mit der Audio-CD dargeboten wird, eindeutig zu wenig ist. Zudem werden die Informationen passiv dargeboten; der Zuhörer ist nicht in die Lerneinheit integriert. Seine Aufmerksamkeit wird nicht „gebunden“. Die Wahrscheinlichkeit, dass er nach kurzer Zeit gedanklich abdriftet und damit nur einen Teil des Lernstoffes aufnimmt, ist sehr groß und der Lernerfolg daher relativ gering.

Von einem interaktiven Hilfesystem erwartet man, dass es die Aufmerksamkeit des Lernenden bindet, indem es ihn in die Lerneinheiten integriert. Man erhofft sich davon, dass der Lernerfolg durch das Prinzip *learning by doing* erhöht wird. Durch schnelle Lernerfolge könnten Berührungsängste gegenüber einem Sprachbediensystem abgebaut werden.

2 Ein Tutorial mit passiven und interaktiven Anteilen

Es sollte also ein Lernmaterial entwickelt werden, das passive als auch interaktive Anteile beinhaltet. Die Idee war hier, die Vorteile von passiven und interaktiven Lernmaterialien auszunutzen und die Nachteile zu reduzieren. Mit den passiven Lerneinheiten können z. B. Dinge erklärt werden, die nicht direkt ausprobiert werden können. Zum anderen können mit diesen Lerneinheiten Erstnutzer „abgeholt“ werden, die sich in Ruhe erst einmal das Prinzip eines SBS oder einen Beispieldialog ansehen möchten, ohne selbst in Aktion treten zu müssen. Mit diesen Lerneinheiten soll also verhindert werden, dass der Nutzer bereits zu Beginn erste negative Erfahrungen sammelt. Er wird nicht überfordert und kann damit erste Ängste vor dem SBS abbauen.

Die interaktiven Lerneinheiten können von fortgeschrittenen bzw. technikversierten Anwendern genutzt werden, die den Lernstoff aus den passiven Lerneinheiten selbst ausprobieren möchten. Eine negative Erfahrung mit dem SBS wird dadurch verhindert, dass innerhalb des Tutorials nur wenige Kommandos sprechbar sind, was wiederum eine Steigerung der Erkennleistung garantiert [5].

Ein Nachteil bei einem Tutorial mit passiven und interaktiven Anteilen besteht in einem erhöhten technischen Aufwand. Da das Tutorial in das Sprachbediensystem integriert wird, können die Hardware- und Teile der Software-Ressourcen des SBS genutzt, somit die Informationen multimodal dargeboten und auch interaktive Lerneinheiten realisiert werden. Die Integration des Tutorials in das SBS hat zudem den Vorteil, dass das Lernmaterial für den Nutzer jederzeit im Fahrzeug verfügbar ist und nicht – wie die Audio-CDs zur *Linguatronic* – im Handschuhfach liegen bleibt.

Um dem Nutzer des Tutorials auch ein Lernmaterial an die Hand zu geben, in dem er einzelne Informationen schnell wieder findet, sollten das Tutorial einfach und klar strukturiert sein und die Menüpunkte eindeutige Bezeichnungen haben.

Das Tutorial ist - außer in den interaktiven Übungen - haptisch zu bedienen. Schließlich wird davon ausgegangen, dass sich überwiegend Erstnutzer erst das Bedienprinzip des SBS anschauen möchten. Gleichzeitig ermöglicht die visuelle Ausgabe des Menüs und die haptische Bedienung eine bessere Orientierung innerhalb des Tutorials. Die visuelle Ausgabe der Menüpunkte kann schnell vom Nutzer erfasst werden, während der Anwender aufgrund des Memorierungsproblems schnell überfordert wäre, wenn ihm die Menüpunkte nur vorgelesen werden würden.

Anhand dieser Anforderungen ergibt sich die Menüstruktur, wie sie in Abb. 1 dargestellt ist.

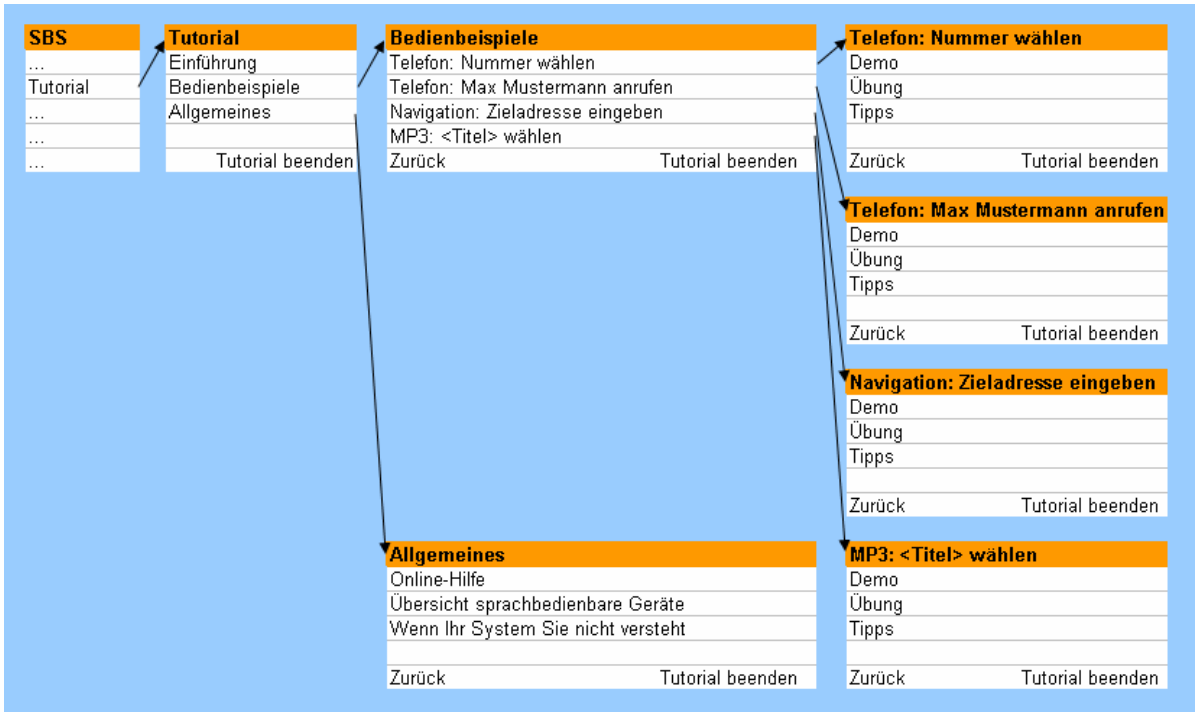

Abbildung 1 - Menüstruktur des Tutorials

Nach Aktivierung des Tutorials gelangt man in die erste Menüebene des Tutorials, das Hauptmenü, in dem zwischen den Punkten *Einführung*, *Bedienbeispiele* und *Allgemeines* gewählt werden kann.

2.1 Einführung

In der Einführung wird dem Anwender gezeigt, dass er das SBS, aktiviert indem er die *Push-to-Talk*-Taste drückt (siehe Abb. 2). Anschließend wird demonstriert, wann er sprechen kann und welche Geräte per Sprache bedienbar sind. Anhand eines kurzen Beispiels wird deutlich gemacht, wie die einzelnen Geräte aufgerufen werden. Da diese Lerneinheit überwiegend für

den Erstnutzer gedacht ist, der sich noch mit dem System selbst vertraut machen muss, ist diese Lerneinheit passiv angelegt.

Abbildung 2 - Drücken der Push-to-Talk-Taste

2.2 Bedienbeispiele

An zweiter Stelle im Hauptmenü befinden sich die Bedienbeispiele. Durch Aktivieren des Buttons gelangt der Nutzer in die nächsttiefere Menüebene, in der er auswählen kann, zu welcher Applikation er ein Bedienbeispiel aufrufen möchte. Es werden vier ausgewählte Themen angeboten, die je nach Verfügbarkeit der Geräte und je nach Hersteller-Wünschen variieren können. Durch den modularen Aufbau können diese einfach ausgetauscht werden. Im dargestellten Tutorial werden die Bedienbeispiele *Telefon: Nummer wählen*, *Telefon: Namen anrufen*, *Navigation: Zieleingabe* und *MP3: Titel wählen* angeboten. Diese vier Beispiele wurden einerseits ausgewählt, da man davon ausgeht, dass sie gegenüber anderen sprachbedienbaren Funktionen häufiger verwendet werden. Andererseits ist bei diesen vier Anwendungen eine Eingabe per Sprache schneller und komfortabler als eine haptische Eingabe, was dem Anwender auch nahegelegt werden soll. Besonders deutlich wird dies bei der Eingabe einer Zieladresse, bei der man mehrere Informationen - eventuell auch buchstabenweise - eingeben muss.

Bei der Auswahl eines der vier Themen gelangt der Nutzer in ein weiteres Menü, in dem er – wie in Abb. 3 zu sehen ist - zwischen einer *Demo*, einer *Übung* und *Tipps* wählen kann. Die Demo präsentiert passiv einen Beispieldialog, der im Anschluss in der interaktiven Übung vom Anwender selbst ausprobiert werden kann. Die ergänzenden *Tipps* werden wiederum passiv präsentiert.

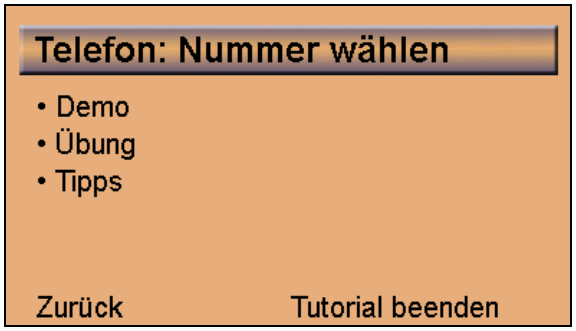

Abbildung 3 - Menü *Telefon: Nummer wählen*

2.3 Allgemeines

Der Menüpunkt *Allgemeines* im Hauptmenü beinhaltet Informationen, die dem fortgeschrittenen Nutzer hilfreich sein können (siehe Abb. 4). In der Lerneinheit Online-Hilfe erfährt der Nutzer Wissenswertes zu der in das SBS integrierten Hilfe. Welche Geräte er per Sprache bedienen kann, kann er sich in einer Übersicht anhören und durchlesen. Diese Übersicht ist zwar bereits in der Einleitung zu finden, doch ein zweiter Zugang an dieser Stelle bietet sich für fortgeschrittene Nutzer an, die bereits ein SBS bedient haben und nur schnell erfahren möchten, welche Geräte per Sprache bedienbar sind. Unter *Das System versteht Sie nicht* gibt es Erläuterungen zur richtigen Sprechweise und zur Geräuschreduktion.

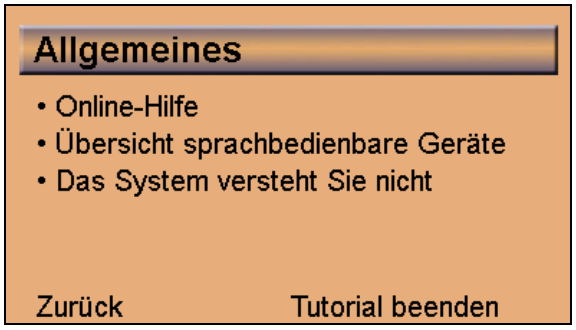

Abbildung 4 - Menü *Allgemeines*

2.4 Navigationsbuttons

Innerhalb des Tutorials stehen dem Anwender verschiedene Navigationsbuttons zur Verfügung. In den Untermenüs sind die Buttons *zurück* und *Tutorial beenden* zu finden. Im Hauptmenü des Tutorials gibt es jedoch nur die Möglichkeit *Tutorial beenden*, was an dieser Stelle dem *zurück* entspricht.

In den Übungen hat der Nutzer bis zum Aktivieren des Dialogs die Möglichkeit, über einen *Beenden*-Button die Übung wieder zu verlassen. Während des Übungsdialogs muss er das Kommando *Abbruch* sprechen, um die Übung vorzeitig zu beenden. Innerhalb der passiven Lerneinheiten wird ein *weiter*-Button angeboten, mit dem man jeweils zum nächsten Thema innerhalb der Lerneinheit springen kann. In der Lerneinheit *Übersicht sprachbedienbare Geräte* steht dem Anwender ein *zurück*-Button zur Verfügung, über den er zurück in das Menü *Allgemeines* gelangt.

2.5 Sprachausgaben

Im Tutorial werden drei verschiedene Stimmen genutzt. Eine Stimme ist die *Tutor*stimme, die den Anwender durch das Tutorial begleitet und in der Einleitung, den Tipps und in den Lerneinheiten des Menüpunktes *Allgemeines* spricht. In den Demos werden für die Beispieldialoge zwei weitere Stimmen genutzt. Eine Stimme ist die *System*-Stimme, die mit der Stimme des Sprachbediensystems übereinstimmen sollte. Die zweite Stimme ist die *Nutzer*-Stimme, die an Stelle des Benutzers die Kommandos spricht. In den Übungen wird ebenfalls die *System*-Stimme verwendet. Die Rolle der *Nutzer*-Stimme übernimmt der Lernende selbst.

Während der Designphase gab es viele Diskussionen zu den Inhalten und dem Umfang der Sprachausgaben. Die Texte dürfen nicht zu lang sein. Sie müssen aber ausführlich genug sein, um die wichtigen Inhalte zu vermitteln. Es ist auch fraglich, wie umfangreich die Sprachausgaben in den Menüs sein sollten. Fühlen sich die Anwender abgeholt, wenn Sprachausgaben die Menüs begleiten oder sollte davon ausgegangen werden, dass die

Bezeichnungen der Menüpunkte so selbsterklärend sind, dass nur in den Lerneinheiten selbst Sprachausgaben dargeboten werden sollten?

Für dieses Tutorial wurde ein Mittelweg gewählt. Dabei wurde auf Sprachausgaben, die nur die Display-Ausgaben wiederholen würden, verzichtet. In den passiven Lerneinheiten gibt es keine einleitenden Worte. Nur in den interaktiven Übungen gibt es kurze Einführungssätze, in denen dem Anwender wichtige Informationen, wie z. B. das Start-Kommando für den Dialog, mitgeteilt werden. Am Ende einer jeden Lerneinheit ertönt ein Ende-Sound, der dem Lernenden verdeutlichen soll, dass die Lerneinheit erfolgreich beendet wurde. Damit wird dem Anwender ein positives Feedback gegeben, unabhängig davon ob er eine Übung sehr gut bewältigt hat oder ob er mehrere Versuche benötigt hat.

2.6 Realisierung

In seiner aktuellen Version ist das Tutorial integriert in das *VoiceCar*-Demo-System von *Harman/Becker*. Es wurde auf einem PC implementiert und ist dort auch lauffähig. Der Sprachdialog ist in JSGF (=**J**ava **S**peech **G**rammar **F**ormat) und der firmeninternen Sprache GDML (=**G**eneric **D**ialog **M**odelling **L**anguage) implementiert. Die graphische Komponente wurde in XML (=E**x**tensible **M**arkup **L**anguage) und GDML programmiert.

3 Evaluierung des Tutorials

Das Tutorial wurde in seiner aktuellen Version von 6 Testpersonen evaluiert. Diese Tests sollten dazu genutzt werden, grobe Fehler aufzudecken. Daher konnten durchaus auch Testpersonen gewählt werden, die nicht der Zielgruppe des Tutorials entsprechen. Alle sechs Versuchspersonen, drei Frauen und drei Männer, sind technikversiert und als *Harman/Becker*-Angestellte in einem technischen Beruf tätig. Bei der Auswahl der Probanden wurde darauf geachtet, dass es sich um neu eingestellte Mitarbeiter handelte, die keine oder kaum Erfahrungen mit SBS haben. Vier der Probanden waren zwischen 18 und 30 Jahre alt und zwei zwischen 31 und 50 Jahren. Fünf der Testpersonen gaben an, dass sie keine Hilfsmittel wie Handbücher, Tutorials oder ähnliches nutzen, um ein neues Gerät kennen zu lernen. Sie würden es selbst ausprobieren.

Die Probanden bekamen 8 Aufgaben gestellt, die sie in einer Testumgebung am PC lösen sollten. Diese Aufgaben sollten u. a. Aufschluss darüber geben, ob das Bedienprinzip des SBS verständlich dargestellt wird, die Menüstruktur klar und die Bezeichnungen der Menüpunkte sowie der Lerneinheiten eindeutig sind und dem Anwender helfen, Informationen schnell wieder zu finden. Die Tests sollten auch zeigen, ob die Aufteilung in passive und interaktive Lerneinheiten sinnvoll ist.

Die Akzeptanztests erbrachten insgesamt ein positives Feedback. Die Tester fühlten sich vom Tutorial nicht überfordert. Die übersichtliche Menüstruktur sowie gleichförmig formulierte Sprachausgaben in ähnlichen Menüs halfen dem Nutzer, sich im Tutorial zu orientieren und Informationen schnell wiederzufinden. Die Aufteilung der Bedienbeispiele in *Demo*, *Übung* und *Tipps*, also in *passive* und *interaktive* Lerneinheiten sowie die Verwendung von drei verschiedenen Stimmen zur Verdeutlichung der Dialoge band die Aufmerksamkeit der Probanden und wurde positiv angenommen. Die begleitenden Sprachausgaben in den Menüs wurden als hilfreich angesehen und ebenfalls positiv bewertet. Die Formulierungen der Lerninhalte waren eindeutig. Der Ende-Sound am Ende jeder Lerneinheit wurde als positiv empfunden. Auf die Frage, ob die Tester an Stelle des Sounds ein Lob oder dergleichen erwarten würden, antworteten sie, dass sie den Ende-Sound gegenüber einem Lob bevorzugen würden.

Als Verbesserungsvorschläge wurde angemerkt, dass die Tutorstimme schneller sprechen sollte und deren Texte kürzer sein könnten. Die Navigationsbuttons könnten allgemein durch Icons ergänzt werden, und innerhalb des Tutorials sollte es einen jederzeit verfügbaren *Exit*-Button geben, nicht nur in den Menüs.

4 Zusammenfassung

Mit dem Tutorial wurde ein geeignetes Lernmaterial entwickelt, welches für Erstnutzer ebenso geeignet ist wie für fortgeschrittene SBS-Anwender. Passive Lerneinheiten vermitteln die Grundlagen und geben dem Novizen einen Einstieg in die Nutzung des SBS ohne ihn zu überfordern. Interaktive Lerneinheiten binden die Aufmerksamkeit des Lernenden und helfen ihm, sich schnell durch *learning by doing* mit dem Bedienprinzip des SBS vertraut zu machen und die erforderlichen Kommandos zu lernen. Mit Hilfe einer klaren Menüstruktur, ähnlich aufgebauten Menüs sowie eindeutigen Bezeichnungen von Menüpunkten und Lerneinheiten kann sich der Anwender innerhalb des Tutorials leicht orientieren.

Das Tutorial könnte problemlos erweitert werden. Beispielsweise könnten mit Hilfe einer History hinter den erfolgreich beendeten Lerneinheiten Häkchen gesetzt werden, um dem Anwender eine Hilfe zu geben, welche Lerneinheiten er bereits besucht hat.

Da davon ausgegangen wird, dass überwiegend Erstnutzer das Tutorial verwenden, um einen Einstieg in die Nutzung eines SBS zu finden, wurde das Tutorial so implementiert, dass es im Wesentlichen haptisch zu bedienen ist. Bisher ist es nur möglich in den Übungen eine Spracheingabe zu tätigen. In einer weiteren Entwicklungsrunde könnte das Tutorial auch so erweitert werden, dass der bereits fortgeschrittene Nutzer auch per Sprache durch das Menü des Tutorials navigieren kann.

Literatur

[1] Koerner, M.: Speech Recognition – The future now! Prentice Hall, 1196

[2] Höhne, F.: Entwurf und Implementierung eines interaktiven Tutorials für ein sprachgesteuertes Radio- und Navigationssystem. Masterarbeit an der BTU-Cottbus in Zusammenarbeit mit Harman/Becker Automotive Systems Ulm, 2007

[3] Fellbaum, K.: Mensch Maschine Kommunikation. Vorlesungsscript SS 2006, Kapitel 1

[4] Salmen, A.: Multimodale Menüausgaben im Fahrzeug. Herbert Utz Verlag 2002, p. 16

[5] Gibbon, D.: Handbook of Standards and Ressources for Spoken Language Systems. Mouton de Gruyter-Verlag, 1997