EVALUIERUNG VON LEXIKALISCHER ADAPTION BEI SPRACHDIALOGEN IM FAHRZEUG

Jakob Landesberger[1], *Ute Ehrlich*[2] *und André Berton*[2]

[1]*Universität Ulm,* [2]*Daimler AG*
jakob.landesberger@gmx.de

Kurzfassung: Die Anpassung an die Wortwahl des Gesprächspartners ist in der zwischenmenschlichen Kommunikation ein wichtiger Faktor für einen erfolgreichen Dialog. In der sprachlichen Interaktion mit einem Computer ist dieses Phänomen ebenfalls feststellbar. Zoltan-Ford, Brennan und Gustafson beispielsweise bewiesen dies in mehreren Experimenten [16][6][10]. In dieser Arbeit wurde in zwei Studien untersucht, wie der subjektive Eindruck von Probanden variiert, wenn diese sprachlich mit lexikalisch adaptiven und nicht-adaptiven Systemen interagieren. In einer ersten *Wizard-of-Oz*-Studie mit einem allgemein gehaltenen Versuchsaufbau konnte der positive Einfluss von lexikalisch adaptiven Systemen nachgewiesen werden. Die Bewertung der Probanden zeigte, dass die Antwort mit demselben Begriff präferiert wurde. Aufbauend auf diesen Resultaten wurde in einer zweiten Studie evaluiert, ob sich die Ergebnisse auf den Kontext eines Fahrzeugs, bei dem die Hauptaufgabe nicht das Führen eines Dialogs ist, übertragen lassen. Dieses Experiment wurde mit Hilfe eines Prototyps in einem Fahrsimulator durchgeführt. Neben dem Dialogführen mussten die Probanden eine Fahraufgabe lösen, wobei diese den Hauptteil ihrer Aufmerksamkeit forderte. Bei diesem Studiendesign präferierten die Probanden ebenfalls das adaptive System hinsichtlich derselben Aspekte wie in der *Wizard-of-Oz*-Studie. Dabei bestärkte sie zum einen die Anpassung in dem Gefühl zu wissen, wie mit dem System zu reden ist. Zum anderen fühlten sich Nutzer bei der nicht-adaptiven Antwort korrigiert.

1 Einleitung

Späterstens seit Apples Assistenzsystem Siri[1] sind Sprachdialogsysteme im alltäglichen Leben angekommen. Siri ermöglicht dem Nutzer per Spracheingabe eine große Bandbreite an Funktionen am Smartphone auszuführen ohne dafür eine Taste drücken zu müssen. Grundsätzlich bieten solche *Conversational Interfaces* große Vorteile, da die Sprache dem Menschen ein unmittelbar zugängliches Mittel ist [12]. Viele dieser Systeme erwarten jedoch eine zu spezielle Wortwahl und sind wenig anpassungsfähig wodurch eine sichere und intuitive Interaktion nicht gewährleistet werden kann [15].
In der zwischenmenschlichen Kommunikation spielt Anpassung für den Erfolg eines Dialogs eine entscheidende Rolle. In fast jedem alltäglichen Dialog ist Adaption beobachtbar. Dabei nähern sich meist beide Gesprächspartner einander an, man spricht dann von Konvergenz bzw. konvergentem Verhalten [3] [9]. Ein Teil dieser Annäherung ist die Adaption an die Wortwahl des Gegenübers. Diese Anpassung ist auch im Bereich der Interaktion mit einem Computer durch mehrere Experimente belegt.
Zoltan-Ford zum Beispiel zeigte durch ein Experiment, dass sich Nutzer eines Sprachsystems

[1]https://www.apple.com/de/ios/siri/, 01.2016

sehr stark hinsichtlich der Wortwahl und des Satzbaus durch das wiederholte anzeigen einer Fehlermeldung, falls andere Wörter verwendet wurden, beeinflussen lassen [16].
Brennan et al. versuchte Nutzer dahingehend zu beeinflussen, dass diese sich der Wortwahl des Systems anpassen. Diese Adaption gelang durch die Wiederholung des entsprechendes Wortes oder dem direkten Hinweis mit Hilfe einer Fehlermeldung bei gesprochenen Dialogen besonders schnell [6].
Gustafson et al. baute auf dieser Forschung auf und untersuchte in zwei Experimenten, wie stark Nutzer die Wortwahl und den Satzbau aus der Fragestellung in deren Antwort wieder verwenden. Dabei wurde gezeigt, dass die meisten Nutzer sich stark an die Fragen des Systems anpassen und besonders bei natürlicher Sprache große Teile aus der Fragestellung unverändert wiederholen [10].
Die aufgeführten Untersuchungen zeigen, dass der Nutzer eines Sprachdialogsystems sich zum Teil sehr stark an die Wortwahl und die Syntax des Computersystems anpasst. Diese Experimente geben allerdings kaum Aufschluss darüber, ob diese Anpassung für den Nutzer einen erhöhten kognitiven Aufwand bedeutet, oder ob durch die starre Systemreaktion, die vom Nutzer als Belehrung aufgefasst werden könnte, die Präferenz, dieses System zu benutzen, gesenkt wird. Branigan gibt Hinweise, dass sich der Eindruck eines Computersystems mit Hilfe unmittelbare linguistische Adaption, nicht verbessern lässt. Dazu führt sie in einem Experiment aus dem Jahr 2005 auf, dass durch diese Art der Adaption ein Spiraleffekt erzeugt wird. Der Nutzer wird durch die adaptive Antwort des Systems lediglich in der Verwendung eines Begriffs nachdrücklich bestärkt [5] [4].
Durch die Verwendung eines Versuchsaufbaus der die Wortwahl des Nutzers spiegeln kann, soll überprüft werden ob sich ein positiver Effekt auf die Einstellung des Nutzers gegenüber dem Sprachdialogsystem erzielen lässt. Diese Auswirkung könnte durch das Gefühl des sofortigen Verständnisses und die suggerierte richtige Wortwahl ausgelöst werden. Weiterhin müsste das Bewusstsein, wie mit dem System zu sprechen ist, bestärkt werden. Je nach Erwartungshaltung an ein Sprachdialogsystem könnte zusätzlich bewirkt werden, dass im Nutzer durch das Benutzen des vermeintlich richtigen Begriffs ein Erfolgserlebnis ausgelöst wird. Ob dieser positive Effekt auftritt und messbar ist, gilt es mit Hilfe zweier Experimente herauszufinden.
In den Abschnitten 2 und 3 werden die beiden Experimente vorgestellt. Dabei wird zuerst auf die verwendeten Methoden und den Versuchsaufbau eingegangen. Im Anschluss erfolgt die Präsentation und Interpretation der erzielten Ergebnisse. Am Ende dieses Beitrags wird in Abschnitt 4 ein Fazit über alle Resultate gezogen und ein kleiner Ausblick dargeboten.

2 Evaluation lexikalischer Adaption mit Hilfe einer Wizard-of-Oz-Studie

Mit Hilfe des ersten Experiments sollte durch ein offenes Studiendesign grundlegend geklärt werden, ob Unterschiede im subjektiven Eindruck von Nutzern zu Sprachdialogsystemen mit verschiedenen Arten der Systemantwort feststellbar sind.

2.1 Studiendesign

Bei dieser Studie handelte es sich um ein *Wizard-of-Oz*-Experiment. Diese Art der Evaluation wurde gewählt, um systemseitiges Fehlverhalten, welches die Probanden negativ beeinflussen könnte, auszuschließen [7][1]. Die Aufgabenstellung bestand immer aus einer vorgegebenen Frage und einer Zahl, die der Proband durch einen Begriff ersetzte. Dieser Ausdruck konnte gebildet werden, indem auf dem angegebenen Raster (vgl. Abbildung 1) eine Bezeichnung für das Bild mit der passenden Zahl gefunden wurde. Die so vervollständigte Frage stellte der Nutzer an das System. Diese grafische Form der Aufgabenstellung wurde gewählt, damit der

Proband in seiner Wortwahl nicht beeinflusst wird. Entsprechend dem getesteten Systemverhalten entsprach die Systemantwort der Wortwahl des Teilnehmers *(identisch)*, bestand aus einem Überbegriff *(allgemein)*, oder es wurde ein gleichwertiges Synonym als Bezeichnung gewählt *(gegenteilig)*. Diese Variationen in der Systemantwort bezogen sich nur auf den Begriff, welcher vom Nutzer anhand des Bildes festgelegt wurde.
Der Begriff unabhängige oder exogene Variable beschreibt einen Parameter, der in einer Studie gezielt verändert wird. Diese Veränderung soll Einfluss auf die abhängige Variable haben. Die abhängige oder endogene Variable entspricht dem Parameter, der gemessen wird. Die Veränderungsstufen der unabhängigen Variable wird als Ausprägung bezeichnet [2] [8].
Somit handelte es sich bei der unabhängigen Variable in diesem Experiment um die Art der Systemantwort. Diese Stellgröße hatte drei Ausprägungen: *identisch*, *gegenteilig* und *allgemein*. Die Auswirkungen dieser Manipulation auf die endogene Variable, die Präferenz des Nutzers, wurde gemessen. Diese Messgröße setzte sich aus den subjektiven Eindrücken des Probanden in Form von Antworten aus Fragebogen und einem Interview zusammen.
Für die Erfassung der abhängigen Variablen wurden zwei Methoden angewandt. Zum Einen wurde ein Fragebogen entworfen, welcher nach dem Beenden jedes inhaltlichen Blocks von dem Probanden ausgefüllt wurde. Zum Anderen klärte ein teilstandardisiertes Interview zum Abschluss des Experiments verbleibende Fragen und gewährte nähere Einblicke in die Ansichten der Interviewteten. Als Grundlage des Fragebogens wurde der *Subjective Assessment of Speech System Interfaces (SASSI)* verwendet [11].
Um bei der Auswertung Reihenfolgeeffekte auszuschließen, wurden die Probanden zufällig in Gruppen eingeteilt, welche sich durch eine unterschiedliche Ausführungsreihenfolge der Aufgaben auszeichneten. Aus Gründen der besseren Durchführbarkeit werden nicht alle durch Permutation erzeugten Reihenfolgen unterschieden, sondern nur drei. Die Aufteilung erfolgt mit Hilfe eines *normalized Latin Square* [2].
Die für diesen Versuch herangezogenen Versuchspersonen bestanden aus 18 Männer und sechs Frauen. Die Altersspannen reichte von 18 bis 44 Jahren. Das durchschnittliche Alter betrug 25,1 Jahre. Weiterhin erfolgte die Erfassung, welche Erfahrung mit Sprachsystemen bereits vorlagen, und wie diese Erfahrungen bewertet wurden. Dabei gaben alle bis auf einen Probanden an, dass sie Sprachsteuerungserfahrung mit mindestens einem Gerät besitzen. Alle Probanden hatten dabei mit demselben Gerät Erfahrung gesammelt, dem Smartphone.
Die gewonnene Praxis beschränkte sich allerdings meist auf eine kurzzeitige, rein experimentelle Nutzung, denn die Frage *„Wie oft benutzen Sie die Sprachsteuerung der oben genannten Geräte“* bewerteten die meisten Probanden mit der niedrigsten Punktzahl. Die Frage nach der Bewertung der grundlegenden Erfahrung mit Sprachsteuerung wurde im Durchschnitt auf einer Skala von durchwegs frustrierend (1) bis durchwegs positiv (5), mit 2,75 bewertet. Das entspricht einer leicht negativen Tendenz.

Abbildung 1 - Auszug der Aufgabenstellung für die Begriffswahl

2.2 Auswertung

Die Ergebnisse aus dem ersten Experiment wurden entsprechend der drei Blöcke *identisch*, *allgemein* und *gegenteilig* ausgewertet und paarweise verglichen. Zwischen der allgemeingültigen und der *gegenteiligen* Systemantwort gibt es bei keiner Frage einen signifikanten Unterschied $(p = 0,613; p = 0,509; p = 0,472; p = 0,317)$. Somit scheint es für die Versuchspersonen keinen Unterschied zu machen, ob mit einem Synonym oder einem Hyperonym geantwortet wurde. Beides wurde als annähernd gleich schlecht bzw. gut eingestuft.
Auch die Freundlichkeit des Systems *(„Ich empfand das System als freundlich“)* erzielte bei keinem der übrigen Vergleiche eine Signifikanz $(p = 0,228; p = 0,172)$. Diese Erkenntnis wird durch die Aussagen der Probanden unterstützt. Dabei wurde festgestellt, dass die Freundlichkeit eher anhand der Stimme und des Satzbaus festgemacht wurde, beides Parameter, die während des Experiments nicht verändert wurden.
Auf den Dialogerfolg *(„Meine Formulierung hat zur gewünschten Reaktion geführt“)* hatten die unabhängigen Variablen ebenfalls keinen Einfluss $(p = 0,093; p = 0,054)$. Der Dialog galt bei den Probanden als erfolgreich, wenn das System die Frage dem richtigen Gegenstand zugeordnet hatte. Diese Zuordnung wurde seitens der Probanden in jedem Durchgang richtig interpretiert. Auch im Falle einer Antwort mit einem Synonym war den Versuchspersonen klar, dass sich das System auf den vom Sprecher ausgewählten Gegenstand bezieht.
Frage 2 *(„Ich hatte das Gefühl, das System wollte mich korrigieren “* $p = 0,000; p = 0,000$*)*

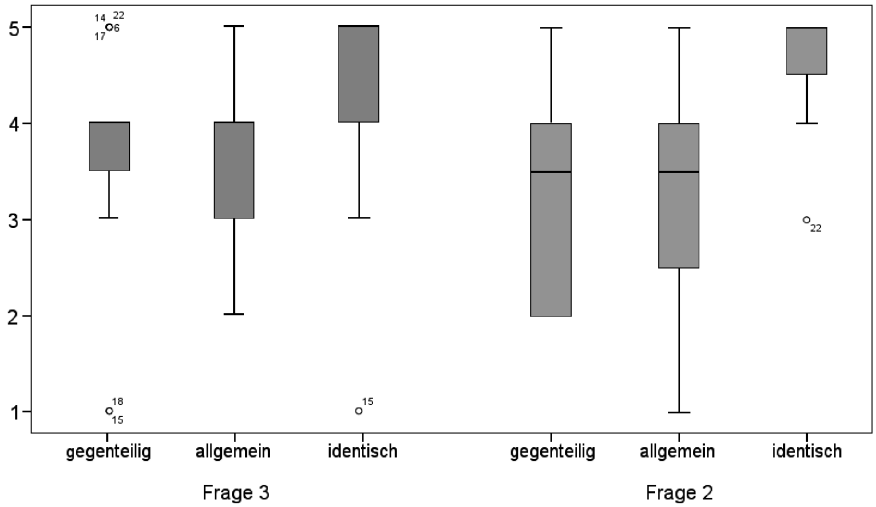

Abbildung 2 - Auswertung der Frage 2 und Frage 3

und Frage 3 *(„Ich hatte das Gefühl, ich habe genau das gesagt, was das System hören wollte “* $p = 0,005; p = 0,003$*)* zeigten bei dem jeweiligen Vergleich der *gegenteiligen* und *allgemeinen* mit der *identischen* Systemantwort signifikante Unterschiede. Die Ausprägung der Unterschiede sind in einer normalisierten Darstellung in Abbildung 2 zu sehen. Anhand der Grafik kann erkannt werden, dass die Probanden sowohl die Antwort mit einem Hyperonym als auch die mit einem Synonym als wesentlich korrigierender empfanden als die mit demselben Begriff. Es wird außerdem deutlich, dass durch die Spiegelung der Wortwahl der Nutzer das Gefühl erzeugt wurde, sie hätten genau den Begriff gewählt, den das System versteht. Im Gegensatz dazu schneidet die *gegenteilige* und *allgemeine* Antwort deutlich schlechter ab.
Die Ergebnisse der qualitativen Erhebungen unterstützen die quantitativen Daten. 14 von 24

Probanden gaben an, die Antwort mit demselben Begriff sei das beste System. Acht fanden alle drei Systeme gleich gut, bzw. konnten sich für keinen Favoriten entscheiden, und nur zwei werteten die *allgemeine* Antwort auf den vordersten Platz. Keiner favorisierte explizit die *gegenteilige* Antwort. Hier gibt es einen deutlichen Unterschied zwischen den Probanden, welche die Systeme differenzieren konnten (zehn von zwölf Versuchspersonen favorisieren das *identische*) zu denen, welchen nichts auffiel (vier von zwölf fanden das adaptive System am besten). Die Resultate belegen, dass die Probanden in diesem Kontext die Verwendung der eigenen Wortwahl in der Systemantwort gegenüber einem Synonym oder Hyperonym bevorzugen. Diese Bevorzugung drückt sich durch ein Gefühl der geringeren Korrektur und dem erhöhten Wissen, wie mit dem System zu sprechen ist, aus.

3 Evaluation lexikalischer Adaption im Kontext eines Fahrzeugs

Der zweite Versuch sollte belegen, dass ein adaptives Sprachdialogsystem im Kontext eines Fahrzeugs hinsichtlich des Eindrucks der Nutzer besser abschneidet als ein statisches System. Dabei war der Dialog die Nebenaufgabe, und die Versuchsperson musste sich in erster Linie auf das Steuern des Fahrzeugs konzentrieren.

3.1 Studiendesign

Für diesen Zweck wurde ein Prototyp zur adaptiven Sprachsteuerung einer Fahrzeug-Klimaanlage entwickelt. Damit ein Fehlerverhalten des Prototyps den Versuch nicht gefährdet, wurden die möglichen Funktionen sehr stark eingegrenzt. Folgende Funktionen können per Sprache aktiviert werden:

- Lüftung erhöhen / reduzieren,
- Temperatur erhöhen / reduzieren,
- Klimaanlage ein- / ausschalten.

Dem Probanden wurden zwei Dialoge präsentiert. Bei einem antwortete das System adaptiv, indem es die Wortwahl des Nutzers verwendete. Dabei wurden sowohl das Objekt und Hauptverb als auch eventuell auftretende Hilfsverben wiederverwendet. Bei dem zweiten gab es eine statische Antwort. Während der Dialoge musste der Nutzer eine Fahraufgabe lösen. Somit wurde der Aspekt des sprachlichen Aufgabenlösens zur Nebentätigkeit. Der Versuch fand innerhalb eines Fahrsimulators statt. Der Innenraum entsprach dabei dem eines echten Autos. Beides ist essenziell, um den Kontext des Fahrzeuges real darzustellen.
Diese Studie misst welches von zwei Sprachdialogsystemen die Probanden präferierten. Ent-

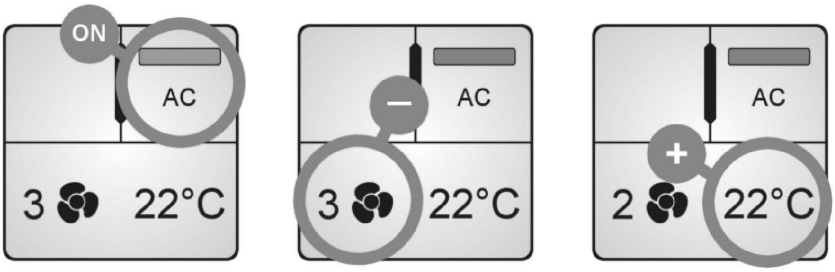

Abbildung 3 - Auszug der Aufgabenstellungen zur Klimaanlagensteuerung

scheidend für den Erfolg der Untersuchung war, genau wie bei der *Wizard-of-Oz*-Studie, dass die Probanden unbeeinflusste Sprachkommandos abgaben. Somit wurden auch hier als Art der Aufgabenstellung Bilder gewählt. Diese sind auszugsweise in Abbildung 3 zu sehen.
Damit ein Lerneffekt das Experiment nicht ungewollt beeinflusste, wurden die Probanden zufällig in zwei Gruppen unterteilt [14]. Je nach Gruppe starteten die Probanden mit der statischen oder der adaptiven Antwort.
Neben dem Sprachdialog bestand die Aufgabe des Nutzers im Bewältigen einer Fahraufgabe. Hierfür wurde der *Contre Task* gewählt. Die Fahraufgabe bestand darin, dass mithilfe des Lenkrads stetig auf ein eingeblendetes Ziel zugesteuert werden sollte. Über dem Ziel war eine Ampel eingeblendet; wenn diese rot aufleuchtete, musste die Bremse, wenn diese grün aufleuchtete, musste das Gaspedal betätigt werden [13].
Um die Eindrücke der Versuchspersonen zu erfassen, wurde ein Fragebogen und ein Interview entwickelt. Der Fragebogen besteht aus zwölf Items, die meisten davon entspringen dem *SASSI* [11].
Die entstandene Stichprobe bestand aus 21 Probanden, wobei es keine Überschneidung mit der Stichprobe der *Wizard-of-Oz*-Studie gibt. Das Alter der Probanden betrug zwischen 21 und 35 Jahren. 19 von 21 Personen hatten bereits vor dem Experiment mindestens einmal Erfahrungen mit Sprachsteuerung gesammelt. Diese beschränkte sich meist auf das Ausprobieren der Funktion. So wurde im Durchschnitt angegeben, dass eine Nutzung von Sprachdialogsystemen maximal ein bis zweimal im Monat stattfindet. Den persönlichen Eindruck legten die Versuchspersonen, auf einer Skala von eins bis fünf, auf drei fest. Das entspricht einer neutralen Erwartungshaltung.

3.2 Auswertung

Abbildung 4 zeigt das Ergebnis des *Gesamtscores* des Fragebogens. Dabei wird deutlich, dass das adaptive System positiver eingestuft wird. Das Resultat beschreibt einen stark signifikanten Unterschied zwischen den beiden Ausprägungen *adaptiv* und *statisch* ($p = 0,02$). Die Aussagekraft des *Gesamtscores* ist bei dieser Studie jedoch nicht hoch, da nur Teile des *SASSI*-Fragebogens verwendet wurden.
Frage 6, 11 und 12 wurden als Kernfragen herausgearbeitet und im folgenden näher betrachtet. Hier ist zu beachten, dass auf Grund des wiederholten Testens über derselben Stichprobe das Signifikanzniveau nach unten korrigiert werden muss ($\alpha' = \frac{\alpha}{3}$). Trotz der Korrektur wurde das adaptive System bei Frage sechs *(„Es ist klar, wie man mit dem System zu sprechen hat.“* $p = 0,005$*)* und Frage elf *(„Ich hatte das Gefühl, das System wollte mich korrigieren.“* $p = 0,003$*)* signifikant besser bewertet. Frage zwölf *(„Ich hatte das Gefühl, ich habe genau das gesagt, was das System hören wollte.“)* erreicht keine Signifikanz ($p = 0,023$).
Die qualitative Befragung bestätigt sowohl die positiven Ergebnisse aus der quantitativen Untersuchung, als auch die Resultate aus der *Wizard-of-Oz*-Studie. Bemerkten die Probanden, dass eines der Systeme sich an die Wortwahl anpasste, so favorisierten sie klar das adaptive System. Die angepasste Antwort des Prototypen wurde als *„persönlicher“*, *„freundlicher“* und *„besser“* bezeichnet. Eine Versuchsperson merkte in Bezug auf die statische Einstellung an: *„Irritierend, wenn das System mit einem anderen Begriff antwortet, als den, den man selbst benutzt hat. “*. Diese Aussage ist durchaus erstaunlich, da in der Regel alle gängigen Sprachdialogsysteme eine statische Antwort verwenden.
War es den Probanden nicht möglich festzustellen, dass das eine System sich an die Wortwahl angepasst hat, verglichen sie die verwendeten Begriffe. In Folge dieses Vergleich schnitt das adaptive System besser ab. So wurden Begriffe aus der statischen Antwort als *„unüblich“* oder *„hochgestochen“* bezeichnet.

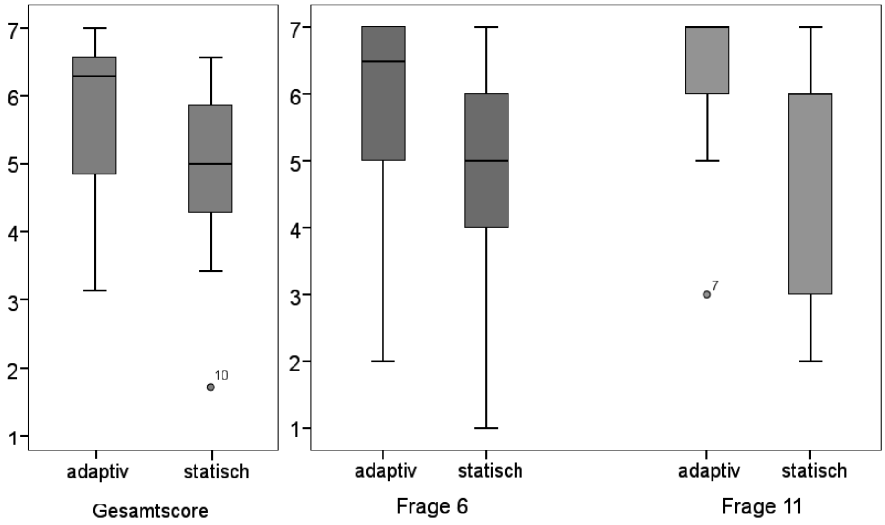

Abbildung 4 - Auswertung des Gesamtscore, der Frage 6 und Frage 11

Diese negative Haltung von Probanden gegenüber den verwendeten Begriffen im statischen System kann ein Zeichen dafür sein, dass im Vergleich zur eigenen Wortwahl ein anderer Begriff als unpassendere Bezeichnung empfunden wird. Dies würde wiederum den Wert adaptiver Sprachdialogsysteme unterstreichen. Eine andere mögliche Erklärung könnte sein, dass der verwendete Begriff allgemein als unüblich gilt. Für die erste Erklärung spricht, dass nur ein sehr geringer Anteil der Probanden die Wortwahl kritisiert hat. Um dies allerdings gänzlich zu klären, wären weitere Untersuchungen nötig.
Der durchgeführte Versuch zeigt, dass die Ergebnisse der *Wizard-of-Oz*-Studie auch im Kontext eines Fahrzeugs gelten. Eine adaptive Antwort hinsichtlich der Wortwahl des Probanden zeigt auf mehrere Weisen einen positiven Effekt. Zum einen wird das Gefühl der Korrektur, welches bei einer starren Antwort, die nicht die vom Nutzer verwendeten Begriffe enthält, deutlich reduziert. Zum anderen erhöht sich bei einer angepassten Antwort die Klarheit, wie mit dem System zu reden ist.

4 Fazit

Der lexikalischen Adaption auf Seiten eines Computers konnten bei einem Dialog mit Probanden erfolgreich positive Eigenschaften nachgewiesen werden: Es kann festgehalten werden, dass junge Menschen um die 25 Jahre mit hinreichender Erfahrung bei der Sprachbedienung, Sprachdialogsysteme mit lexikalischer Adaption gegenüber statischen in zwei Punkten favorisieren. Zum einen bestärkt sie die Anpassung in dem Gefühl zu wissen, wie mit dem System zu reden ist, und zum anderen fühlen sie sich durch die statische Antwort stärker korrigiert. Dieses Ergebnis konnte sowohl in dem Kontext eines Fahrzeuges bei dem der Dialog die Nebenaufgabe war, als auch im Rahmen eines Versuchs ohne Ablenkung erzielt werden. Bemerkenswert ist, dass bei beiden Versuchen diejenigen Probanden, die die adaptive Funktionsweise erkannten, diese klar favorisierten. Somit könnte ein Versuch mit längeren Dialogen oder über einen größeren Zeitraum Ergebnisse hervorbringen, die noch stärker für ein adaptives System sprechen. Gleichzeitig müsste gewährleistet sein, dass unter anderem die Adaption von umgangssprachlichen Ausdrücken verhindert wird. Diese lexikalische Anpassung könnte zu einer gegenteiligen

Reaktion beim Nutzer führen, da die gespiegelte Wortwahl als Fehler oder sogar als verbale Parodie interpretiert werden könnte. In wie weit dies für den Nutzereindruck relevant ist, ist Teil weiterer Forschungen.

Literatur

[1] BERNSEN, N. O., H. DYBKJÆR und L. DYBKJÆR: *Designing interactive speech systems: From first ideas to user testing*. Springer Science & Business Media, 1998.

[2] BORTZ, J. und N. DÖRING: *Forschungsmethoden und evaluation*. Springer-Verlag, 2013.

[3] BRANIGAN, H. P., M. J. PICKERING und A. A. CLELAND: *Syntactic co-ordination in dialogue*. Cognition, 75(2):B13–B25, 2000.

[4] BRANIGAN, H. P., M. J. PICKERING und J. F. MCLEAN: *Priming prepositional-phrase attachment during comprehension.*. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(3):468, 2005.

[5] BRANIGAN, H. P., M. J. PICKERING, J. PEARSON und J. F. MCLEAN: *Linguistic alignment between people and computers*. Journal of Pragmatics, 42(9):2355–2368, 2010.

[6] BRENNAN, S. E.: *Lexical entrainment in spontaneous dialog*. Proceedings of ISSD, 96:41–44, 1996.

[7] DAHLBÄCK, N., A. JÖNSSON und L. AHRENBERG: *Wizard of Oz studies: why and how*. In: *Proceedings of the 1st international conference on Intelligent user interfaces*, S. 193–200. ACM, 1993.

[8] FAHRMEIR, L., R. KÜNSTLER, I. PIGEOT und G. TUTZ: *Statistik*. Springer-Verlag, 2007.

[9] GARROD, S. und G. DOHERTY: *Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions*. Cognition, 53(3):181–215, 1994.

[10] GUSTAFSON, J., A. LARSSON, R. CARLSON und K. HELLMAN: *How do system questions influence lexical choices in user answers?*. In: *Eurospeech*. Citeseer, 1997.

[11] HONE, K. S. und R. GRAHAM: *Towards a tool for the subjective assessment of speech system interfaces (SASSI)*. Natural Language Engineering, 6(3&4):287–303, 2000.

[12] LEMON, O.: *Conversational Interfaces*. Springer, 2012.

[13] MAHR, A., M. FELD, M. M. MONIRI und R. MATH: *The contre (continuous tracking and reaction) task: A flexible approach for assessing driver cognitive workload with high sensitivity*. Adjunct Proceedings of the 4th AutomotiveUI, Portsmouth. ACM, 2012.

[14] NIELSEN, J.: *Usability engineering*. Morgan Kaufmann, 1993.

[15] UNIVERSITY OF UTAH: *Talking to Your Car is Often Distracting*, Nov. 2015.

[16] ZOLTAN-FORD, E.: *How to get people to say and type what computers can understand*. International Journal of Man-Machine Studies, 34(4):527–547, 1991.