

## NUTZERZENTRIERTER ENTWURF VON MULTIMODALEN BEDIENKONZEPTEN

*I. Wendler, A. Jatho, I. Kraljevski, M. Wenzel*

*voice INTER connect GmbH  
ines.wendler@voiceinterconnect.de*

### **Kurzfassung:**

Durch nutzerzentrierte Entwicklung von User Interfaces lässt sich der Nutzer frühzeitig in den Entwicklungsprozess einbeziehen. Bereits bevor ein funktionierendes System existiert, erlaubt der Wizard-of-Oz- (WOz) Test, ein Nutzererlebnis zu simulieren, mit Anwender-Zielgruppen Tests durchzuführen sowie Konzepte und Hypothesen zu prüfen. Der Nutzen für die User Interface-Entwicklung ist zweifach: Durch parallele Datenaufzeichnung werden einerseits Testkorpora gewonnen. Andererseits erlauben die Nutzerexperimente Rückschlüsse auf Nutzerverhalten, -erwartungen und die Gestaltung des Bedienablaufs für ein fehlertolerantes, robustes Verhalten. Der Beitrag dieses Papers liegt darin, die Methodik nutzerzentrierten Designs für multimodale Bedienkonzepte und eine Experimentalanordnung mit konstanten Bedingungen für 13 verschiedene Sprachen darzulegen und zu verifizieren, dass sich ein Dialogdesign mit nur geringfügigen Anpassungen sprachübergreifend einsetzen lässt. In einem aktuellen Projekt wurde in einer Vorbereitungsphase für 16 Sprachvarietäten eine Online-Befragung mit potentiellen Nutzern zur Klärung genereller Erwartungen durchgeführt und um Anregungen für das Dialogdesign zu erhalten. In einer Durchführungsphase wurden Dialoge für Nutzungsszenarien entworfen und implementiert, die möglichst spontansprachliche Äußerungen der Nutzer hervorrufen sollten. Bei der Durchführung des WOz-Tests wurde ein multimodales Korpus aufgenommen, welches Sprach- und Videoaufnahmen sowie Transkriptionen der Probandeninteraktion umfasst. In der abschließenden Evaluationsphase wurde die Zufriedenheit der Probanden mithilfe einer Nachbefragung ermittelt, welche überwiegend positiv ausfiel.

### **1 Einleitung**

Nutzerzentriertes Design ist eine Methode zur Entwicklung von Software, welche dazu beiträgt, dass das Produkt den Bedürfnissen der Nutzer entspricht. Gestaltungsentscheidungen werden nicht allein vom Designer der Software getroffen sondern der Nutzer wird in den Mittelpunkt des Entwicklungsprozesses gestellt. vgl. [1]

Die zu entwickelnde Software ist ein Dialogsystem mit Spracherkennung und Sprachsynthese zur Bedienung mehrerer elektronischer Geräte im Smart-Home-Bereich, wie beispielsweise eines Kaffeevollautomaten. Der Nutzer soll bereits zu Beginn von dessen Entwicklung eingebunden werden, auch wenn zu diesem frühen Zeitpunkt noch keinerlei Programmierung vorgenommen worden ist. Daher werden eine Vorabbefragung, Nutzertests mit einem Prototypen in Form eines Wizard-of-Oz- (WOz) Tests sowie eine Nachbefragung durchgeführt.

Menschen passen sich beim Sprechen an ihr jeweiliges Gegenüber an und reagieren unterschiedlich auf verschiedene Personen, ebenso wie auf verschiedene Computersysteme. Um verlässliche Daten aufzunehmen wird daher dem potentiellen Nutzer im Rahmen des WOz-Tests suggeriert, er interagiere mit einem Dialogsystem. vgl. [2]

In der Literatur sind WOz-Tests [3, 4, 5, 7], WOz-Tools [3, 6, 8] für Dialogsysteme auch im Smart-Home-Bereich und teils mit mehreren Sprachen beschrieben. Trotzdem hat sich gezeigt, dass aufgrund der Unterschiedlichkeit der Anwendungen mit Hinblick auf das geplante Zielsystem ein sehr individueller Versuchsaufbau und ein eigenes WOz-Tool nötig sind.

Ebenso ist den Autoren keine Veröffentlichung eines WOz-Tests bekannt, der für eine so große Anzahl von Sprachen entworfen wurde. Das nachfolgende Paper überprüft den Ansatz, dass der Dialogentwurf in einer Sprache nur durch Übersetzung und marginale Anpassungen in mehrere Zielsprachen überführt werden kann und dokumentiert die Vorgehensweise an einem konkreten Beispiel.

## 2 Vorbereitungsphase

Im Vorfeld des eigentlichen WOz-Tests wurde eine Online-Umfrage für 16 Sprachvarietäten durchgeführt, wobei die Ergebnisse von 870 potentiellen Nutzern erhoben wurden. Im Folgenden wird auf die Ergebnisse der deutschen Vorbefragung eingegangen. Entsprechend der definierten Zielgruppe sind alle Teilnehmer der Umfrage über 30 Jahre alt, wobei die Gruppe der 50 bis 59-Jährigen am stärksten vertreten ist. Der Anteil an weiblichen und männlichen Teilnehmern beträgt jeweils exakt 50%.

Die Vorbefragung ermittelt neben allgemeinen Fragen wie Erfahrungen mit der Bedienung von Smart-Home-Devices und Sprachsteuerung sehr umfangreich und detailreich, wie sich die Nutzer die Interaktion mit einem intelligenten User Interface vorstellen und auf welche Art und Weise sie konkrete Parameter in spezifischen Nutzerszenarien eingeben möchten.

Beispielsweise zielte eine zentrale Frage, die für die gesamte Dialoggestaltung des Zielsystems relevant ist, darauf ab, zu ermitteln, welche Länge von Sprachein- und -ausgaben die Nutzer favorisieren: *„Sie möchten ein Getränk aus dem Kaffeevollautomaten beziehen. Wie würden Sie Ihr Anliegen mitteilen? Bitte sortieren Sie die Antwortoptionen nach Ihrer persönlichen Vorliebe (Plätze 1 bis 3).“*

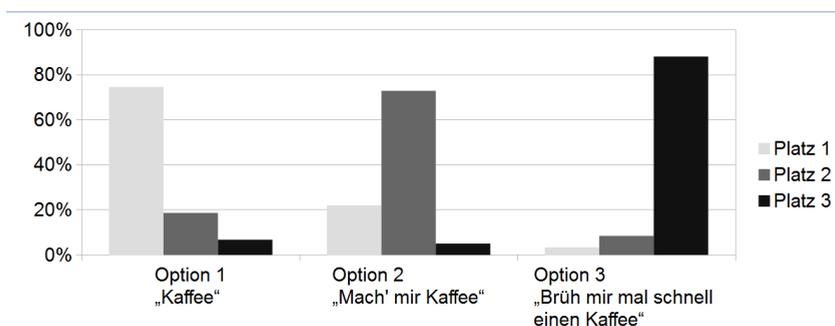


Abbildung 1 – Antworten auf die Frage nach der Befehlslänge in der Vorbefragung (Deutsch)

Es wurde mit 75% eindeutig die sehr knapp formulierte Option 1 auf den ersten Platz gewählt. Diese Tendenz zur Kürze bestätigte sich auch in weiteren Fragen. Dementsprechend wurden für den WOz-Test Prompts implementiert, die so knapp wie möglich gehalten wurden, ohne dabei unhöflich zu klingen und ohne die Intonation negativ zu beeinflussen. Beispielsweise führte die Übersetzung des Prompts „Welches Gerät?“ ins Englische („Which appliance?“) zu einer Intonation, die von den Probanden als Aussage wahrgenommen wurde und nicht als eine Frage.

Um sprachliche Nutzereingaben von nicht an das System gerichteten Äußerungen zu trennen, spielen ein robust erkennbares Schlüsselwort für die Systemaktivierung und ein für den Nutzer intuitives Schlüsselwort zur Korrektur von Eingaben eine wichtige Rolle. Die Ergebnisse der deutschen Vorbefragung legen nahe, dass das im WOz-Test vorgegebene Schlüsselwort „Korrektur“ von vielen Nutzern intuitiv verwendet wird. Situationsbezogen kommen auch Synonyme wie „falsch“, „stop“, „ändern“ etc. zum Einsatz.

Weitere Fragen ermittelten unter anderem die Nutzerpräferenzen für das Aufrufen der Hilfefunktion, das Ansprechen einzelner Geräte, die Dialoggestaltung nach erfolgreicher Parameterabfrage und die Anzahl der Parameter, welche innerhalb eines Turns eingegeben werden. Implizit wurden weitere Synonyme für Geräte- und Parameterbezeichner erfasst.

### **3 Ausführungsphase**

#### **3.1 Implementierung der Dialoge**

Auf Basis der Vorabefragung wurden die Dialoge für das simulierte Zielsystem zunächst auf Deutsch in das WOz-Tool implementiert. Die Prompts wurden später in 12 weitere Sprachen übersetzt und in die bestehende Dialogstruktur importiert. Das war mit wenigen geringfügigen Anpassungen möglich. Diese Änderungen waren weitestgehend auf Übersetzungsungenauigkeiten zurückzuführen und darauf, dass die TTS durch eine untypische Prosodie teilweise schwer verständlich war.

#### **3.2 Aufgabenstellungen**

Während der Sprachaufnahmen spielten die Probanden 19 unterschiedlich komplexe Nutzerszenarien durch, die ihnen vom Versuchsleiter mündlich als Aufgabe gestellt wurden und erhielten dafür als Gedächtnisstütze je eine Hilfekarte. Die Hilfekarten enthielten Bilder und kurze Schlagworte um den Einfluss auf die Wortwahl der Probanden zu minimieren. Die Probanden wurden dazu angehalten, alle Spracheingaben möglichst ohne Benutzung der Hilfekarten zu tätigen, um ein freies Sprechen hervorzurufen.

Die Aufgabenstellungen orientierten sich an zuvor durch Nutzerstudien ermittelten Nutzerszenarien und decken verschiedene Parameter mehrerer Geräte ab. Beispielsweise lautete eine Aufgabe: „Die Kaffeemaschine arbeitet vollautomatisch. Sie brauchen nur bestellen und Ihr Getränk wird sofort zubereitet. Bitten Sie die Kaffeemaschine um einen Latte Macchiato, einen Kaffee oder einen Espresso“. Auf der zugehörigen Hilfekarte waren lediglich das Gerät und die zu nutzende Funktion angegeben.

Die Schwierigkeit der Aufgaben wurde im Laufe des Tests gesteigert durch komplexere Dialoge, eine steigende Anzahl von einzugebenden Parametern und geplante Fehlreaktionen des Dialogsystems. Für etwa 20% der einzugebenden Parameter war eine Fehlreaktion des Systems eingeplant, um ausreichend Korrekturfälle zur Untersuchung von Korrekturstrategien der Probanden zu provozieren. Die Fehler stammten aus den Kategorien „Einfügung“, „Löschung“ und „Ersetzung“. Während bei der Einfügung ein Bestätigungsprompt auf eine nicht geäußerte Probendenaussage gespielt wurde, wurde bei der Löschung die Eingabe des Nutzers ignoriert bis hin zum Timeout und bei der Ersetzung eine falsche Ausprägung des eingegebenen Parameters bestätigt.

#### **3.3 Versuchsaufbau**

Abb. 1 zeigt schematisch den Aufbau des Smart Home-Labors und des WOz-Kontrollraums. Um die Interaktion so natürlich wie möglich zu gestalten und eine möglichst realitätsnahe Geräuschkulisse zu schaffen, werden Requisiten wie beispielsweise Tassen und Kaffee für die Bedienung der Kaffeemaschine bereitgestellt. Ein technischer Assistent, der Wizard, steuert indessen die Simulation des User Interfaces in einem getrennten Raum. Tragbare Mikrofone

machen ein freies Bewegen im Raum möglich und tragen somit zu einer hohen Immersion in die Nutzungssituation bei. Neben den aufgenommenen Sprachdaten von Versuchsleiter und Proband werden ebenso parallele Videoaufnahmen der Versuche gemacht, welche gespeichert und zum Wizard zur Beobachtung der Szenarien übertragen werden.

Der Wizard startet mit Hilfe des WOz-Tools die Dialogsteuerung der aktuellen Aufgabe. Ein Tablet ermöglicht die Kommunikation zwischen Wizard und Versuchsleiter. Der Proband löst die ihm vom Versuchsleiter übertragene Aufgabe in Interaktion mit den Geräten und Requisiten. Der Wizard simuliert das User Interface, indem er die Eingaben des Probanden durch Drücken von Aktionsfeldern beantwortet. Dadurch werden die Sprachausgabe von vorgefertigten oder variablen Prompts auf Lautsprecher im Testraum, bzw. Geräteaktionen ausgelöst. Die mit dem WOz-Tool ausgeführten Bedienaktionen werden zeitlich protokolliert und gemeinsam mit den Sprachdaten aufgezeichnet. Diese dienen später als Marker für den Test der Dialogsteuerung.

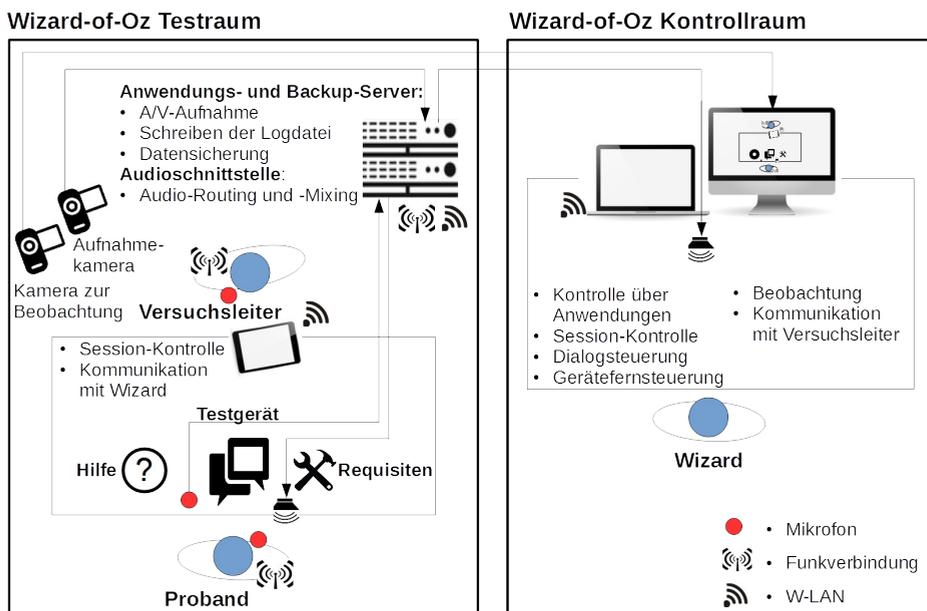


Abbildung 2 - Räumliche Darstellung des Versuchsaufbaus in Draufsicht

Für die Ausgabe der deutschen Prompts wurde eine weibliche TTS-Stimme des Nuance Vocalizer verwendet. Für die weiteren Sprachen wurden je nach Verfügbarkeit Stimmen des Nuance Vocalizer oder von Microsoft verwendet. Die TTS musste für jede Sprache durch Einfügen von Sonderzeichen in den Prompts optimiert werden, da die Prosodie durch die sehr kurzen Prompts nicht optimal war und teils das Verständnis deutlich beeinträchtigt hat.

### 3.4 Wizard und Versuchsleiter

Wizard, Versuchsleiter und Probanden waren jeweils Muttersprachler der Zielsprachen. Während als Wizard insbesondere technikaffine Mitarbeiter akquiriert wurden, die schnell und intuitiv mit dem WOz-Tool umgehen konnten, wurden die Versuchsleiter nach sozialer und sprachlicher Kompetenz beim Umgang mit den Probanden ausgewählt.

### 3.5 Probanden

Zielgruppe waren Muttersprachler der Sprachen Deutsch, Englisch, Französisch, Spanisch, Niederländisch, Italienisch, Finnisch, Dänisch, Norwegisch, Schwedisch, Russisch, Türkisch und Mandarin. Es wurde eine Gleichverteilung männlicher und weiblicher Probanden angestrebt.

Die Probanden unterzeichneten eine Einverständniserklärung für die Nutzung der aufgenommenen Daten und wurden darüber aufgeklärt, dass ihre Aufnahmen anonymisiert gespeichert werden.

Für Deutsch und Englisch wurden je 40 Probanden aufgenommen, in allen weiteren Sprachen galt als Zielstellung, 20 Testpersonen zu gewinnen. Über eine Dauer von acht Monaten nahmen 252 Probanden am WOZ-Test teil.

## 4 Evaluationsphase

### 4.1 Abschlussfragebogen

Anschließend an die Sprachaufnahmen wurde den Probanden ein knapper Abschlussfragebogen präsentiert. Hierin wurde unter anderem überprüft wie leicht den Nutzern der Umgang mit der Sprachbedienung fällt, ob die Prompts verständlich waren und wie sie die Länge der Prompts empfanden.

Auf einer 5-stufigen Skala von „sehr einfach“ bis „sehr anstrengend“ empfand die Mehrheit der der Nutzer die Sprachsteuerung als „einfach“ zu bedienen. Aus Abb. 3a, b geht hervor, dass die Sprachbedienung nicht nur als gleich einfach in den übersetzten Sprachen gewertet wurde sondern, dass sogar mehr der nicht-deutschen Probanden die Bedienung „sehr einfach“ fanden als die deutschen Teilnehmer. Daraus wird gefolgert, dass die Übertragung der deutschen Dialoge in die weiteren Sprachen keinen negativen Einfluss auf die Bedienung des Systems hat. Dies deckt sich ebenso mit den schriftlichen Berichten der Versuchsleiter, in denen nur wenige Abbrüche von Nutzerszenarien dokumentiert sind.

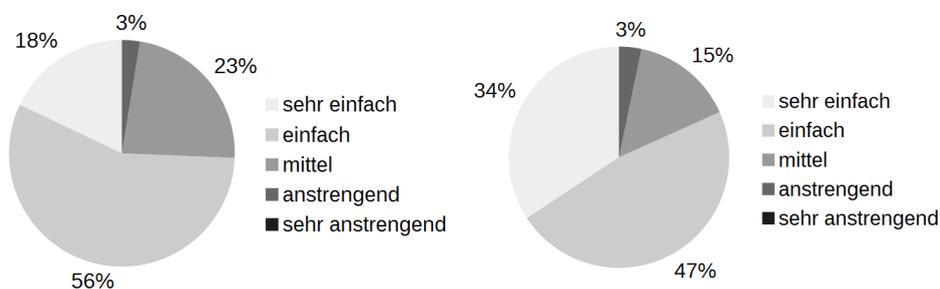
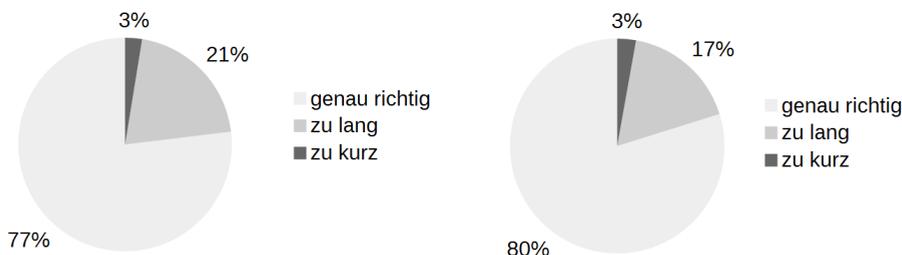


Abbildung 3a, b – Antworten auf die Frage nach der Einfachheit der Bedienung der Sprachsteuerung in der Nachbefragung a) für Deutsch und b) für alle anderen Testsprachen

Weiterhin wurde in der Nachbefragung geprüft, ob die Probanden die allen Nutzerszenarien zugrunde liegende Dialogstruktur verstanden hatten. Dies war bei fast allen Probanden der Fall und wird als Hinweis darauf aufgefasst, dass die aus dem Deutschen übernommene Dialogstruktur auch für Sprecher anderer Sprachen nachvollziehbar ist.

Die Ergebnisse der Vorbefragung in Bezug auf die Promptlänge bestätigten sich auch in der Nachbefragung wieder. In Folge der Bemühungen, die Prompts kurz und knapp zu

gestalten, empfanden 77% der Probanden die Promptlänge als genau richtig. Weitere 21% der Befragten hingegen fanden die Prompts dennoch zu lang, vgl. Abb. 4a. Wie aus Abb. 4b hervorgeht, ist auch nach der Übertragung der Prompts in weitere Testsprachen im Mittel keine Verschlechterung der Nutzerzufriedenheit in Bezug auf die Promptlänge zu erkennen.



**Abbildung 4a, b** – Antworten auf die Frage nach der Promptlänge in der Nachbefragung a) für Deutsch und b) alle anderen Testsprachen

Im Anschluss an die Nachbefragung wurden die Probanden darüber aufgeklärt, dass sie während des Tests mit dem Wizard interagiert haben. Kein Proband hatte den Verdacht geäußert, nicht mit einem Dialogsystem, sondern mit einem Menschen interagiert zu haben.

Im Nachgang zu jeder Aufnahmeeinheit wurde von jedem Aufnahmeteam ein Bericht angefertigt, in welchem Schwachstellen bei der TTS, in der Übersetzung und Dialogführung festgehalten und bei Bedarf vor der nächsten Einheit korrigiert wurden.

## 4.2 Aufbereitung der Sprachkorpora

Schließlich wurden die Aussagen aller Probanden orthografisch transkribiert, nach Störgeräuschen, Verständlichkeit, Zögern bzw. Pausen und dialektalen Aussprachevarianten markiert und gleichzeitig auf Turn-Ebene segmentiert. Die Transkriptionsarbeit wurde von je einem Versuchsleiter oder Wizard geleistet. Diese wurden zur Sicherstellung einer einheitlichen Vorgehensweise ausführlich eingewiesen. Gemäß den Anforderungen des Projektes wurden die Aussprachevarianten nicht im SAMPA-Standard [9] sondern orthografisch notiert. Verwendet wurde das Tool Subtitle Edit [10], welches einfacher zu bedienen ist als gängige Transkriptionsprogramme wie z. B. Praat. Durch den Import der Prompts aus dem Logfile in das Oszillogramm bzw. Spektrogramm wurde das Transkribieren vereinfacht und beschleunigt. Die beim Segmentieren festgesetzten und in eine Textdatei gespeicherten Zeitpunkte für Beginn und Ende der Turns wurden für das automatisierte Schneiden der Aufnahmen mit einem Batch-Processing-Skript verwendet. Die so entstandenen Audiodateien und ihre Transkriptionen werden als Testdaten für den Spracherkenner des Zielsystems verwendet.

## 5 Daten

In der Vorabbefragung erhobene Daten zu den Bedürfnissen von potentiellen Nutzern flossen bereits in das Design des WOz-Tests ein. Das in der Nachbefragung erhobene Nutzerfeedback sowie die Berichte der Aufnahmeleiter fließen in die Entwicklung des Zielsystems ein.

Das im WOz-Test erstellte Korpus besteht aus Audio- und Video-Rohdaten mit einem Umfang von etwa 125 Stunden. Darin enthalten sind etwa 4.800 Dialoge über alle Sprachen hinweg. Die Audioaufnahmen der Probanden weisen eine Abtastrate von 16 kHz und eine

Abtasttiefe von 16 bit auf. Die Signal-to-Noise-Ratio der deutschen Audiodaten liegt zwischen 0,95dB und 42,49dB und im Durchschnitt bei 24,24dB. Die Rohdaten aus den deutschen Aufnahmen umfassen 1004 Dateien (Mikrofone, zwei Kameras, Metadaten), welche einen Speicherplatz von 295 GB einnehmen.

**Tabelle 1** – Merkmale des deutschen Teils des Korpus

Dauer der segmentierten Aufnahmen	3h 17min 30sec
Dauer der sprachlichen Äußerungen	1h 49m 59s
Anzahl von Dialogen	ca. 740
Mindestanzahl von Turns eines einzelnen Sprechers über alle 19 Nutzerszenarien	78
Maximale Zahl von Turns eines einzelnen Sprechers über alle 19 Nutzerszenarien	224
Durchschnittliche Anzahl von Turns eines einzelnen Sprechers über alle 19 Nutzerszenarien	145
Mindestanzahl von Turns für ein Nutzungsszenario	2
Maximale Anzahl von Turns für ein Nutzungsszenario	21
Durchschnittliche Anzahl von Turns für ein Nutzungsszenario	7,6
Minimale Dauer eines Turns	0,13s
Maximale Dauer eines Turns	15,878s

Tabelle 1 listet einige Merkmale des deutschen Korpus auf. Eine Auswertung der Wortzahl pro Äußerung ergab, dass die in Tabelle 1 genannten Turns überwiegend aus sehr knappen Befehlen mit einer Länge von ein (39,05%) oder zwei (36,90%) Wörtern bestanden, vgl. Abb. 5. Im Durchschnitt enthielt ein Turn lediglich 2,3 Worte.



**Abbildung 5** – Auftretenshäufigkeit der Äußerungslängen 1 bis 29 (Deutsch)

Der im Deutschen verwendete Wortschatz (13.290 Wörter insgesamt) und das verwendete Vokabular (601 unterschiedliche Wörter) ergeben eine niedrige Type-Token-Relation von 0,045. Zusammen mit der Auswertung der Äußerungslängen und den Ergebnissen der Vorabbefragung zeichnet sich also ab, dass die Nutzer eine kurze und prägnante Interaktion mit dem Dialogsystem ausgeschmückten Formulierungen vorziehen.

## 6 Zusammenfassung und Diskussion

Bei dem erfolgreich in 13 bzw. 16 Sprachen durchgeführten Experiment mit Vorab- und Nachbefragung wurden Erkenntnisse über die Nutzervorlieben in Bezug auf die Interaktion mit einem Sprach-User Interface, insbesondere die Interaktionslänge, gewonnen.

Es wurde verifiziert, dass die Übertragung eines Dialogdesigns für die Interaktion mit Smart Home-Geräten lediglich durch Übersetzung und mit nur minimalen Anpassungen möglich ist. Es hat sich gezeigt, dass es dem überwiegenden Teil der Nutzer leicht oder sehr leicht fällt, das Dialogsystem zu nutzen und dass eine Übertragung der Dialogstruktur lediglich durch Übersetzung und geringfügige Anpassungen ein mindestens gleich gutes Nutzererlebnis hervorrufen kann.

Eine Einschränkung der Erkenntnisse aus dem gewonnenen Datenmaterial besteht darin, dass für einige Sprachen, beispielsweise Schwedisch, Norwegisch und Finnisch weniger Versuchspersonen zur Verfügung standen, als anvisiert wurde. Aus diesem Grund sind Alters- und Geschlechtsverteilung der Probanden über die Sprachen hinweg nicht identisch.

Da Muttersprachler als Versuchsleiter und Wizard eingesetzt wurden, mussten zwangsläufig verschiedene Mitarbeiter die Aufnahmen durchführen. Trotz intensiver Schulung und Vorbereitung kann daher nicht ausgeschlossen werden, dass die Probanden durch die leicht unterschiedliche Arbeitsweise der Mitarbeiter in Ihrer Interaktion mit dem System und in ihrer Ausdrucksweise beeinflusst wurden.

Das erstellte multimodale Korpus bestehend aus Audio-, Video- und Transkriptionsdaten wird für weiterführende Untersuchungen zu sprachlichen Unterschieden in Bezug auf Dialogführung, Korrekturstrategien sowie phonetische Analysen genutzt.

## Literatur

- [1] LOWDERMILK, T.: *User-centered design*. Sebastopol, CA: O'Reilly Media 2013, S. 6
- [2] DAHLBÄCK, N., JÖNSSON, A., AHRENBORG, L.: *Wizard of Oz studies: why and how*. In: *Proceedings of the 1st international conference on Intelligent user interfaces*. ACM, 1993, S. 2 ff.
- [3] KREBBER, J., MÖLLER, S., PEGAM, R., JEKOSCH, U., MELICHAR, M., & RAJMAN, M.: *Wizard-of-Oz tests for a dialog system in smart homes*. In: *Proceedings of the joint congress CFA/DAGA*, Strasbourg, France, 2004.
- [4] BRUTTI, A., CRISTOFORRETTI, L., KELLERMANN, W., MARQUARDT, L., & OMOLOGO, M.: *WOZ acoustic data collection for interactive TV*. In: *Language Resources and Evaluation*, 44(3), 2010, 205-219.
- [5] LI, A. X., & BONNER, J. V.: *Designing intelligent domestic applications using wizard of oz methodology*, 2011
- [6] SCHLÖGL, S., LUZ, S., & DOHERTY, G. *WEBWOZ: A Platform for Designing and Conducting Web-based Wizard of Oz Experiments*, In: *Proceedings of the SIGDIAL 2013 Conference*, Metz, France, Association for Computational Linguistics, August, 2013, pp. 160–162.
- [7] PETRIK, Stefan. *Wizard of Oz Experiments on Speech Dialogue Systems. Design and Realisation with a New Integrated Simulation Environment*. Masters. Graz University of Technology, Graz. Institute of Signal Processing and Speech Communication, 2004.
- [8] MUNTEANU, Cosmin; BOLDEA, Marian. *MDWOZ: A Wizard of Oz Environment for Dialog Systems Development*. In: *LREC*. 2000.
- [9] WELLS, J.C., *SAMPA computer readable phonetic alphabet*. In: Gibbon, D., Moore, R. and Winski, R. (Hrsg.): *Handbook of Standards and Resources for Spoken Language Systems*. Berlin and New York: Mouton de Gruyter. Part IV, section B, 1997.
- [10] Subtitle Edit, <http://www.nikse.dk/subtitleedit/>, Zugriff 9.1.2016