

MurderMystery: Interaktive Krimigeschichten als Alexa-Skill

Frederik Wollatz, Ina Rentemeister, Valeria Ebinger

University of Duisburg-Essen, Germany
Frederik.Wollatz@stud.uni-due.de
Ina.Rentemeister@stud.uni-due.de
Valeria.Ebinger@stud.uni-due.de

Abstract

Im Rahmen des Praxisprojekts "Praxisprojekt Alexa, hilf mir beim Sprachenlernen" des Lehrstuhls Language Technology Lab der Universität Duisburg-Essen wurde eine Alexa Anwendung entwickelt die mithilfe einer interaktiven Krimi-Geschichte das Hörverstehen sowie das Sprechen auf Englisch trainiert. Die Interaktivität findet ausschließlich in Form eines Dialogs statt. In diesem Paper wird sowohl die Kernidee der Anwendung ausführlich beschrieben als auch der zentrale Lernaspekt, den diese Anwendung dem User bringt. Des Weiteren wurde eine User Study zur Erfassung der Usability der Anwendung durchgeführt. Es lässt sich sagen, dass die Usability der Anwendung bereits in einem guten Bereich liegt, sodass die Anwendung ohne große Probleme genutzt werden kann. Um die Usability weiterhin zu verbessern, wird eine stetige Verbesserung erfolgen, wie zum Beispiel an den Regular Expressions. Die Anwendung nutzt ausschließlich den Hörsinn der User, da weitere Studien zeigen, dass solche Anwendungen den gewünschten Effekt auf das Lernen von Fremdsprachen haben (Nold and Rossa, 2007)

1 Einleitung

Die beschriebene Anwendung MurderMystery ist eine interaktive Krimi-Geschichte, die dem User helfen soll, die englische Sprache zu lernen. Es wird das Kontextverständnis als auch der Wortschatz verbessert. Zunächst werden verwandte Arbeiten beschrieben, sodass die Anwendung in den Kontext gesetzt werden kann. Daraufhin wird die Anwendung ausführlich beschrieben, sowie die Kernidee und der Lernaspekt dargestellt. Um einen Einblick in die Anwendung zu bekommen, werden Beispiele der Zeugenbefragung und des Tutorials dargestellt. Um die Usability der Anwendung zu testen,

wurde eine User Study mit 15 Versuchspersonen durchgeführt. Die Usability wurde mit Hilfe von zwei bewährten Usability-Fragebögen getestet. Darunter fielen der System-Usability-Scale-Fragebogen (Brooke, 2013) sowie der ISONORM 9241/110 Fragebogen von Prümper und Anft (Figl, 2009). Weiterhin wurden selbstentwickelte, anwendungsspezifische Fragen gestellt, um die Anwendung besser einordnen zu können und gegebenenfalls Schwachstellen aufzudecken. Weiterhin wird beschrieben, wie eine umfassende Evaluationsstudie aussähe, die aufgrund von Zeitmangels in diesem Projekt nicht möglich war. Danach folgt die Darstellung der Ergebnisse und die Auswertung der dargestellten Ergebnisse. Im letzten Teil des Papers wird diskutiert, welche Ziele bereits erreicht wurden und welche Möglichkeiten der Erweiterung der Anwendung es noch gibt.

2 Verwandte Arbeiten

Als bereits bestehende Anwendung für interaktive Geschichten lässt sich Earreality aufzählen. Die Erfinder nehmen bereits bestehende Geschichten von anderen Autoren und erstellen daraus interaktive Geschichten. Die deutsche Anwendung lässt sich über eine Webseite (www.ear-reality.de) aufrufen. Dort lassen sich dann Geschichten nach vier Kriterien aussuchen. Nach Genre, Folgen, Spieldauer und Verfügbarkeit. Die Geschichten sind über Alexa verfügbar. So wie Earreality funktioniert, sollte die optimierte Version von MurderMystery auch funktionieren, allerdings wird diese ergänzt mit dem Hauptaspekt der Anwendung: dem Lernen der englischen Sprache, dadurch dass ausschließlich Englisch gesprochen wird. Hierzu wird dann noch die Schwierigkeitsstufe hinzugefügt, nach dem jeweiligen Sprachlevel. Diese kann der Nutzer dann

gezielt auswählen. Auch Rafael Miranda, General Manager bei Neusta Mobile Solutions Berlin, hat einen Post im Alexa Blog verfasst, wo er berichtet, dass interaktive Geschichten mit Alexa ein Comeback haben (Miranda, 2019). Hiermit bezieht er sich auf Earreality. Mit dem Skill „Der Eiserne Falke“ gewannen EarReality sogar den Hauptpreis für einen deutschsprachigen Skill in der Alexa Skills Challenge: Games. Die meisten Geschichten, die bei Earreality ausgewählt werden können, lassen sich dem Genre Fantasy zuordnen. MurderMystery grenzt sich insofern ab, als dass ausschließlich Kriminalgeschichten verwendet werden.

Kriminalgeschichten wurden verwendet, da die Spannung den Leser bei Kriminalromanen mehr motiviert, die Geschichte zu Ende zu lesen (Nusser, 2009). Hier also übertragbar auf den Skill MurderMystery: Durch die interaktive Geschichte und die Spannung des Kriminalfalls werden User über längere Zeiträume die Anwendung nutzen und somit schneller ihre Sprachfertigkeiten im Englischen verbessern. Auch nicht zuletzt dadurch, dass der Stellenwert von Kriminalromanen in der Gesellschaft in allen Schichten sehr hoch. So dass das Interesse für die Applikation aufgrund der Kriminalromane höher als bei anderen Genres sein könnte (Drobez, 2014). Auch das interaktive Storytelling schafft Möglichkeiten zur Entwicklung der Sprache, sozialer sowie kognitiver Fähigkeiten, dadurch dass der User sich aktiv in das Geschehen einbringen muss und nicht ausschließlich zuhören (Goldman, 1998).

Es ist bekannt, das Hörverstehen eine Schlüsselrolle beim Erlernen von Fremdsprachen spielt, vor allem bei jungen Lernenden (Anderson and Lynch, 1988). Die Anwendung nutzt ausschließlich den Hörsinn der User, da weitere Studien zeigen, dass solche Anwendungen einen positiven Effekt auf das Lernen von Fremdsprachen haben (Nold and Rossa, 2007). Außerdem hat das Hören von Geschichten einen positiven Lerneffekt, wenn es um das Vokabeltraining geht (Elley, 1989).

3 Anwendungsbeschreibung

Durch die Applikation "MurderMystery" soll vorrangig das Sprachverständnis und Hörverständnis trainiert werden. Es muss nachhaltig verstanden werden, was gesagt wurde, um den Krimi erfolgreich zu lösen. Ebenfalls sollen neben dem Training von Vokabelwissen, auch

komplexe Geschichten verknüpft werden. An Fertigkeiten sollen Dialogorientiertheit und Formulierungsfähigkeit vorhanden sein sowie Vokabelwissen des Levels B1 oder höher um die Anwendung problemlos zu nutzen. Das Hörverstehen soll trainiert werden, sowie komplexe Sachzusammenhänge erkannt und verstanden werden, sodass das Lösen des Kriminalfalls möglich ist. Des Weiteren werden sowohl Dialoge trainiert als auch das logische Denken der User angeregt sowie gefordert und gefördert.

Es wird durchgehend in englischer Sprache erzählt, gesprochen und nachgefragt. Es gibt während des Spiels durch die Interaktion mit verschiedenen Figuren Akzente, die von schottisch über britisch bis amerikanisch reichen, jedoch einfach zu verstehen sind und auf das Hörverständnis der User keinen Einfluss haben. Das Hörverstehen wird sogar gefördert, da Menschen im echten Leben auch über unterschiedliche Akzente, abhängig von ihren Herkunftsländern, verfügen beziehungsweise die User diesen Akzenten dann begegnen. Daher ist es wichtig auch die Unterscheidung der verschiedenen Akzente zu erlernen (Gramley, 1976).

MurderMystery richtet sich an Englisch Lernende mit jeglicher Muttersprache, da die Anwendung ausschließlich auf Englisch abläuft. Allgemein ist B1 und höhere Level als erforderliches Sprachniveau angesetzt. Das Alter ab dem die Anwendung nutzbar ist, kann nicht eindeutig konkretisiert werden. Allerdings wird die Schwierigkeit der Aufgabenstellung auch am logischen Verständnis festgemacht. Laut dem Schweizer Psychologen Jean Piaget (1999) gliedert sich die Denkentwicklung in vier Stufen, sodass die Anwendung erst ab circa 12 Jahren genutzt werden kann, wenn die 3. Stufe, die "Stufe des logischen Denkens (7-12 Jahre)" (Kesselring, 1999) abgeschlossen ist. Level B1 beinhaltet laut dem europäischen Referenzrahmen, dass die Kernaussagen von allgemeinen Themen auf der Arbeit, in der Schule oder auf Reisen verstanden werden sollen. Weiterhin das Erkennen von Zusammenhängen und bei deutlicher Sprache auch Details. Außerdem sollen faktische Texte zu Interessensgebieten der User verstanden werden. Die Mindestzeit um alle Texte zu hören beträgt 30 Minuten, allerdings ist es zum Lösen des Falls nicht zwingend notwendig alle Texte zu hören. Die Spielzeit beträgt durchschnittlich 15 bis 30

Minuten und ist abhängig vom Sprachniveau, dem logischen Denkvermögen, aber auch der Strategie des Nutzers.

4 Kernidee

Die Kernidee der Anwendung ist es, spielerisch Englisch zu lernen. Dies geschieht durch das Lösen eines Krimis, durch viel Interaktion des Users mit der Anwendung beziehungsweise den Charakteren. Der entwickelte Krimi findet in Wales in einer alten verlassenen Villa statt. Es gibt einen hohen Storytelling Anteil, um das Erlebnis so echt wie möglich zu gestalten (Soleimani and Akbari, 2013).

Zu Beginn wird ein Tutorial angeboten, welches die Spielregeln erklärt, sodass neue User einen einfachen Einstieg finden. Erfahrene User können das Tutorial überspringen. Es gibt insgesamt acht Räume, in denen der User sich bewegen kann. Es ist jederzeit möglich von allen Räumen in jeden Raum zu gelangen. In jedem Raum kann man einen Zeugen befragen, den Raum durchsuchen oder den Raum wieder verlassen und einen anderen Raum aufsuchen. Von jedem Raum aus ist es möglich das Rätsel zu lösen. Des Weiteren ist jeder Raum für die Interaktion identisch aufgebaut, das heißt, dass immer ein Zeuge in einem Raum ist, der Raum auf Anfrage durchsucht sowie verlassen werden kann.

Watson ist der Helfer in der Applikation und dient zur Unterstützung der User bei Nachfragen. Er kann jederzeit durch das Keyword "Watson" angefragt werden. Der User erhält auf Anfrage Zusammenfassungen der bereits betretenen Räume und gesammelten Zeugenaussagen. Ebenfalls kann Watson gefragt werden, welche Räume schon betreten wurden, sodass eine bessere Orientierung für den Nutzer gegeben ist. Außerdem kann er innerhalb von Befragungen Beispielfragen an die Zeugen formulieren, um die Befragung für den Nutzer einfacher zu gestalten.

Der Krimi kann zu jeder Zeit durch den User gelöst werden. Es ist jedoch insgesamt nur drei Mal möglich einen Lösungsvorschlag abzugeben oder einen Tatverdächtigen zu nennen. Wenn der richtige Mörder genannt wurde, wird das Rätsel vollkommen aufgelöst und die Anwendung danach beendet. Wenn der falsche Mörder genannt wurde, kann man weiterhin Zeugen befragen oder Räume durchsuchen. Nennt man allerdings drei Mal den falschen Mörder, flieht der

richtige Mörder. Hierbei wird das Rätsel auch nicht aufgelöst und das Spiel gilt als verloren. Anschließend wird die Anwendung beendet.

4.1 Beispiele

Beispiel für eine Raumdurchsuchung anhand des ersten Raums.

Erzähler: "The Entrance Hall is dimly lit and feels calm and empty. The eerie silence is only disturbed by the echo of your steps on the cold marble floor and the big doors closing behind you. You see several doors and a grande staircase. Next to the staircase are two knights armor on each side. A butler is standing next to one of the doors. What do you want to do in the entrance hall? Would you like to speak to the witness, search the room or leave the room?"

User: "I want to search the room."

Erzähler: "You take a closer look around. Nothing seems unusual to you on the first glance. Everything is perfectly cleaned and there are no traces of an intruder. The murderer must be someone within the house. Then you see it. One of the knights armors positioned next to the staircase is missing it's sword, which could be a murder weapon. What do you want to do in the entrance hall? Would you like to speak to the witness, search the room or leave the room?"

User: "I want to speak to the witness."

Erzähler: "The Butler of the house is a grim looking lanky man with pronounced eyebrows and a hooked nose reaching almost down to his lips. He is wearing a worn down, but meticulously cleaned suit."

Watson: "What do you want to ask this witness? If you need some example questions, ask me."

User: "Example questions, please."

Watson: "You can ask the witness the following questions: Where have you been when the murder was committed? What was your relationship with the lord? Did something seem weird to you lately? Do you have any idea who killed the lord? Could you imagine others committing such a crime? Do you want to ask anything else?"

User: "Where have you been when the murder was committed?"

James: "Polishing silver cutlery with the Maid. Why are you asking? Do I seem like a murderer? I have been working here since a long, long time. I'm becoming older, don't want to risk or lose my job, understand detective?"

Watson: "Do you want to ask anything else?"

User: "What was your relationship with the lord?"

James: "Young and stupid man, you know. He still had a lot to learn. We didn't really have a connection. I was just doing my job here man."

Watson: "Do you want to ask anything else?"

User: "Did something seem weird to you lately?"

James: "A lot of things seem weird these days, don't you think? The Old Lady has acted strangely... Bought a lot of stuff. Go ask her and leave me alone."

Watson: "Do you want to ask anything else?"

User: "Do you have any idea who killed the lord?"

James: "Detective, I'm not gonna tell you the secrets of this house. Never have, never will."

Watson: "Do you want to ask anything else?"

User: "Could you imagine others committing such a crime?"

James: "I was with the maid. Don't know anything about the others, mind my own business, you know."

Watson: "Do you want to ask anything else?"

User: "No, thank you."

Erzähler: "What do you want to do in the entrance hall? Would you like to speak to the witness, search the room or leave the room?"

User: "Leave the room."

Erzähler: "Where do you want to go: The garden? The kitchen? The lounge? The basement? The Servantbedroom? The bedchamber or the library?"

4.2 Technische Umsetzung

Im folgenden Abschnitt betrachten wir, wie die vorgenommenen Ziele Programmiertechnisch umgesetzt wurden. Wir beginnen mit dem generellen Ablauf (Figure 1) und betrachten dann die Details.

Beim Starten des Skills werden die beiden Funktionen "onStart()" und "onLaunch()" aktiviert. Sie setzen den Skill in den Ursprungszustand zurück und erstellen einen String namens "text", welcher zu diesem Zeitpunkt nur "<speak>" enthält. Außerdem erfolgt hier der erste Prompt, also die erste Frage an den User. Dieser Prompt wird zu einem Speechletresponse umgewandelt und von Alexa ausgesprochen. Nun antwortet der User. Diese Antwort wird von Alexa verstanden und in

einen Envelope umgewandelt und an unsere Funktion "onIntent()" weitergegeben. Diese wandelt den Envelope wiederum in einen String namens "userrequest" um. Danach wählt sie den Zustand aus, in dem wir uns gerade befinden. Zustände bestimmen die Verarbeitung von "userrequest". In jedem Zustand wird untersucht, welche Zeit wir in diesem Zustand haben. Beispielsweise hat der Zustand "Zeuge" zwei Zeitpunkte: Der Erste, in welchem der Zeuge geladen und das erste mal beschrieben wird und der Zweite, in welchem der Zeuge befragt wird. Je nach Zeit wird der Funktion "RecUI()" ein anderer Parameter mitgegeben, welcher bestimmt wie die "userrequest" verarbeitet wird. "RecUI()" wandelt die "userrequest" in den Intent des Users um. Dies geschieht auf Grundlage von Regular Expressions. Diese Matchen meist auf einzelne Schlüsselbegriffe die den Willen des Users verraten. Je nach Intent wird nun eine Antwort ausgewählt. Sollte kein Intent erkannt werden, wird die Funktion "pardon()" gestartet, welche als Antwort die Bitte generiert, den Satz umzustrukturieren bzw. zu wiederholen. Die Antwort wird mit der Funktion "speak()" an "text" angefügt. "speak()" umgibt den gesendeten Text hierbei mit den entsprechenden Tags für die Stimme des Sprechers. Zuletzt wird noch die Zeit des Zustandes aktualisiert. Jetzt kommen wir zur letzten Funktion innerhalb von OnIntent: "sende()". Diese hängt "text" noch "</speak>" an und wandelt "text" dann in ein speechletresponse um. Danach wird "text" wieder zurückgesetzt auf "<speak>". Das speechletresponse wird nun von "onIntent()" retourniert und damit von Alexa ausgesprochen. Damit beginnt der Kreislauf von neuem.

Kommen wir noch einmal auf die Zustände zurück. Es gibt fünf Zustände: Das Tutorial, der Raum, der Zeuge, Watson und Solve. Die Zustände sind ineinander geschachtelt, da man immer an seinen Ursprungsort zurückkehren soll (Figure 2). Sollte man aus dem Raum heraus solven wollen, so landet man bei der Rückkehr aus diesem Zustand wieder im Raum, nicht in Watson oder dem Zeugen. Im Tutorial kann man noch nicht den Fall lösen und auch Watson befragen ist noch nicht möglich. Wenn man hingegen in einem Raum ist, kann man diesen durchsuchen, den Zeugen vor Ort befragen, mit Watson reden oder in den nächsten Raum gehen. Vom Zeugen hingegen ist es nicht möglich den Raum zu durchsuchen oder in einen anderen Raum zu gehen. Wenn man

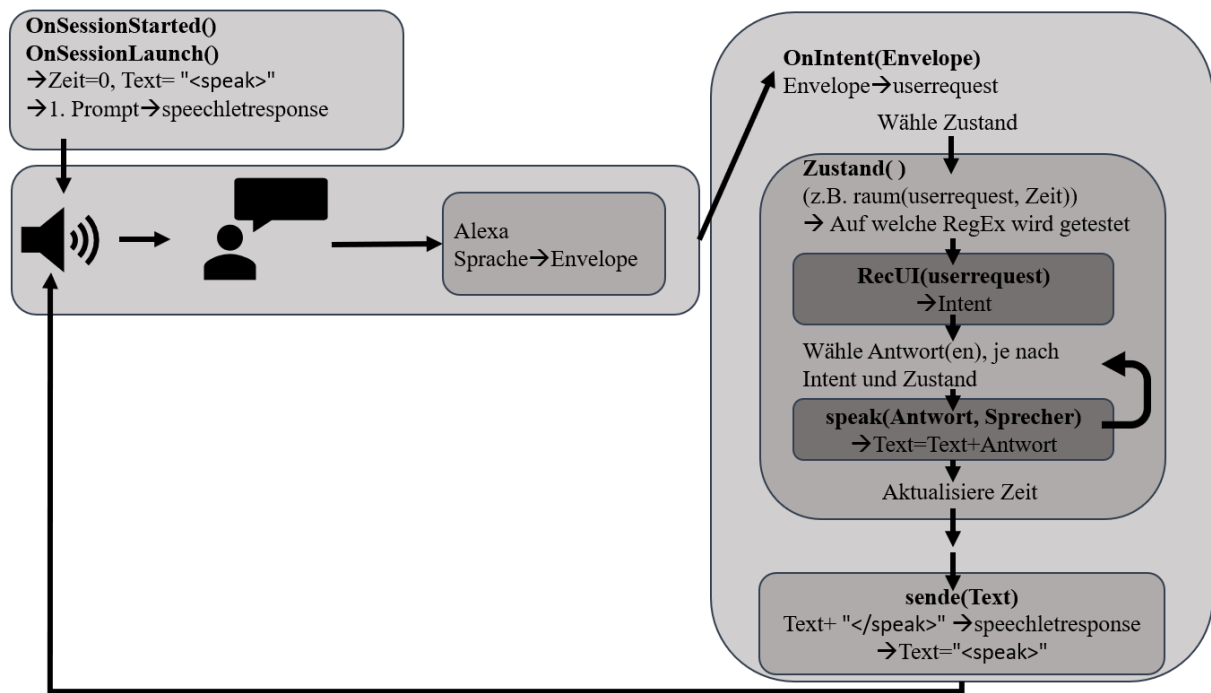


Figure 1: Vereinfachte Darstellung des Programmablaufs.

den Fall richtig löst oder zu häufig falsch liegt, so wird das Spiel beendet.

4.3 Zentraler Lernaspekt

Der zentrale Lernaspekt der angesprochen wird, ist das Verständnis des Kontextes. Da die Anwendung auf Personen des Sprachlevel B1 und höher zugeschnitten ist, wird erwartet, dass bereits viel richtig ausgesprochen wird und für neue Wörter ein entsprechendes Sprachgefühl vorhanden ist, sodass die Bedeutung von unbekannten Wörtern auch aus dem Kontext erschlossen und verstanden werden kann. Sprachliche und grammatische Fehler werden nicht beachtet, da die Anwendung diese nicht erkennt. Die User werden nicht zur Wiederholung der Aussage aufgefordert. Allerdings wird auf die Anfrage einen Raum zu betreten in dem nächsten Satz des Erzählers direkt der Raum nochmals benannt, sodass die User dadurch erkennen können, dass sie es gegebenenfalls falsch ausgesprochen haben, sodass ein Lerneffekt entsteht. Dem Nichtverstehen von einzelnen Wörtern, beziehungsweise ganzen Teilgeschichten, wirken die jeweiligen Zusammenfassungen, die bei Watson abgefragt werden können, entgegen. Außerdem kann der komplette, bis dahin im Laufe des Spiels abgefragte Inhalt, wiederholt angehört werden. Zusätzlich besteht die Option sich ansagen zu lassen in welchen Räu-

men man schon war oder welche Räume man noch nicht besucht hat. Gleiches gilt für die Zeugenbefragungen.

5 User Study

5.1 Methode

Die Studie wurde in zwei Teile gegliedert. Der erste Teil besteht aus einen Interaktionsteil mit der Applikation in welchem der Proband die Applikation testet in dem er das Spiel spielt. Der Zweite Teil der Studie ist zur Erfassung der Nutzererfahrungen, der Usability der Interaktion und der demografischen Daten der Probanden. Der Fragebogen-Teil ist in vier Teile gegliedert. Der erste Teil besteht aus 10 Fragen des System Usability Scale (SUS) Fragebogen (Brooke, 2013). Danach werden selbstentwickelte Fragen gestellt. Es folgt ein Fragebogen ISONORM 9241/110 von Prümper und Anft sowie deskriptive Fragen (Figl, 2009).

Der SUS Fragebogen (System Usability Scale) wird zur quantitativen Erhebung in der Usability-Forschung genutzt (Brooke, 2013). Er wurde 1986 von John Brooke entwickelt. Es ist eine sogenannte "Quick-and-dirty-Methode" die zur Messung der Usability angewandt wird. Dieses ist heutzutage in der Nutzerforschung ein sehr anerkanntes Verfahren. Die Fragen werden auf einer Skala (Likert-Skala von 1 bis 5) von den Nutzern beant-

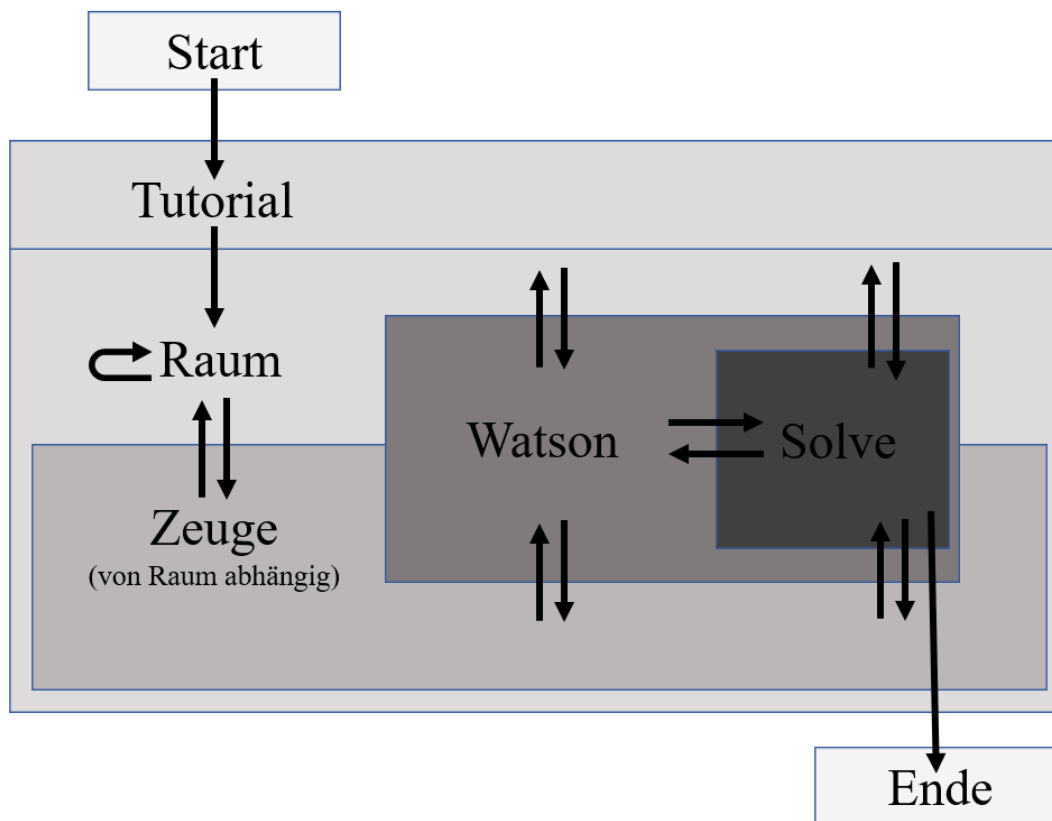


Figure 2: Sämtliche Zustände und der Wechsel zwischen ihnen.

wortet, sodass nachher eine quantitative Auswertung erfolgen kann, die interpretierbar ist. Um die Interpretation zu ermöglichen, wird ein Usability-Wert (SUS-Score) der Applikation gebildet. Hierbei wird der durchschnittliche Score pro Item verwendet. Wichtig hierbei ist, dass die Items 1,3,5,7 und 9 minus 1 genommen werden und die restlichen Items jeweils 5 minus den Itemscore. Die daraus entstehenden Werte werden summiert und anschließend mit 2,5 multipliziert um dann einen Wert zwischen 0 und 100 zu erhalten. Anhand dieses Wertes wird die Nutzerfreundlichkeit einer Software festgemacht werden. Denn je höher der Wert, desto höher die Nutzerfreundlichkeit. (Brooke, 2013). Diese Kategorien werden im Standardverfahren mit Werten von 0 bis 5 kodiert. Die Kodierung ist von der Formulierung abhängig. Bei positiv formulierten Vorgaben erhält die Antwort volle Zustimmung den Wert 5 und eine grundsätzlich ablehnende Antwort den Wert 0. Bei negativ formulierten Vorgaben wird die Antwort volle Zustimmung mit 0 kodiert, eine vollständig ablehnende Antwort erhält den Wert 5 (Brooke, 2013).

Aus den Ergebnissen des SUS-Fragebogens er-

rechnet sich ein sogenannter SUS-Score. Diese Score wird in einer Spanne von 0 bis 100 betrachtet. Während eine Ausprägung von 0 die schlechteste vorstellbare Anwendung beschreibt und 100 die beste vorstellbare Anwendung. Bei der Auswertung des Fragebogens werden die gewonnenen Zahlen addiert – die Summe liegt zwischen 0 und 40 – und anschließend mit 2,5 multipliziert. Aus den Scores der Probanden wird der durchschnittliche Score gebildet. Dieser Durchschnittswert kann nun als Prozentwert interpretiert werden, wobei 100 einem perfekten System ohne Usability-Probleme entsprechen. Werte über 80 Prozent deuten auf eine gute bis exzellente Usability hin. Werte zwischen 60 Prozent und 80 Prozent sind als grenzwertig bis gut zu interpretieren. Werte unter 60 Prozent sind Hinweise auf erhebliche Usability-Probleme.

Es wurden eigene Fragen generiert um genauere Informationen und Bewertungen von den Usern zu der Anwendung zu bekommen. Diese werden ebenfalls auf einer Likert-Skala von 1 bis 5 bewertet. (Stimme gar nicht zu - Stimme voll zu) Ein Beispielitem lautet: "Ich erachte die Anwendung als unterhaltsam". Auch hier wird ein Durch-

schnittswert zur Ermittlung der Nutzerbewertungen und -Erfahrungen ermittelt, indem die einzelnen Mittelwerte nochmal gemittelt werden. Anmerkungen, die von den Nutzern eingefügt wurden, werden einzeln betrachtet.

Als dritter Teil wird der ISONORM 9241/110 von Prümper und Anft (Figl, 2009) genutzt um zu überprüfen, inwieweit die Anwendung die in der gleichnamigen Norm formulierten sieben Gestaltungsanforderungen wie "Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Steuerbarkeit, Erwartungskonformität, Fehlertoleranz, Individualisierbarkeit und Lernförderlichkeit" erfüllt (Figl, 2009). Die Dialogprinzipien werden anhand mehrerer Aussagen bewertet. Die Befragten geben an, ob eher die negative Aussage, die auf der linken Seite steht oder die positive Aussage, die auf der rechten Seite steht, auf die Software zutrifft.

Zum Abschluss der Studie wurden noch deskriptive Werte erhoben. Sowie Alter, Geschlecht, Abschluss, Gaming Erfahrung, und das Englisch Level.

5.2 Umfassende Evaluationsstudie

Eine umfassende Evaluationsstudie würde so aussehen, dass Versuchspersonen randomisiert in mehrere Gruppen eingeteilt werden. Zunächst wird mittels Hörverständnisaufgaben das Sprachniveau der Versuchspersonen ermittelt. Eine Gruppe, die Kontrollgruppe, lernt anschließend während der gesamten Studienlaufzeit gar nicht. Eine andere lernt mit der Anwendung MurderMystery. Je nach Wunsch und Bedarf kann die Anwendung auch mit anderen Lernmöglichkeiten wie Vokabelkarten, Hörverstehen-CDs oder auch Sprachkursen mittels weiterer Gruppen verglichen werden. Zu verschiedenen Messzeitpunkten wird anschließend der Lernerfolg mittels weiteren Hörverständnisaufgaben erfasst werden. Die Anzahl der Probanden hier müsste recht hoch sein um persönliche Abweichungen, wie Lernmotivation, zu minimieren. Außerdem kann der MurderMystery Gruppe im Anschluss ein weiterer Usabilitytest mittels eines Fragebogens vorgelegt werden um so die Anwendung weiter verbessern zu können. Hierfür sollten allerdings fünf Versuchspersonen ausreichen, da es bei jeder weiteren Person unwahrscheinlicher wird, dass neue Usability Information aufgedeckt wird (Turner et al., 2006). Diese Änderungen können mit einer weit-

eren Studie wiederum hinsichtlich ihres Einflusses auf den Lernerfolg untersucht werden.

5.3 Ergebnisse

Insgesamt nahmen 15 Versuchspersonen an der Nutzerstudie teil. Darunter waren 8 weibliche Versuchspersonen und 7 männliche Versuchspersonen. Das durchschnittliche Alter der Versuchspersonen betrug 24 Jahre in einer Range von 18 bis 53 Jahren. Alle Versuchspersonen hatten mindestens Abitur, wovon ein paar über Studienabschlüsse verfügen. Etwa ein Drittel der Versuchspersonen verfügte über Gaming-Erfahrung. Das Sprachniveau der Versuchspersonen erstreckte sich von einem A2-Sprachlevel bis zu einem C1 Sprachlevel.

Es wurden durchschnittlich 1,4 Versuche der Versuchspersonen benötigt, um den Krimi zu lösen, das heißt den Mörder richtig zu benennen. Weiterhin lösten 13 von 15 Personen das Rätsel. Wie sinnvoll die Anwendung zum Sprachenlernen ist, wurde auf einer Skala von 1 bis 5 (1= nicht sinnvoll, 5= sehr sinnvoll) mit einem durchschnittlichen Wert von 4,2 bewertet. Das heißt, die Anwendung wurde als sinnvoll bis sehr sinnvoll eingestuft. Ein Proband beschrieb sein Anwendungserlebnis wie folgt: „Es war sehr interessant und ich kann mir vorstellen, dass man damit gut das Hörverstehen üben kann“. Die Unterhaltsamkeit wurde ebenfalls auf einer Skala von 1 bis 5 (1= nicht sinnvoll, 5= sehr sinnvoll) mit einem durchschnittlichen Wert von 3,33 bewertet. Die Schwierigkeit der Anwendung wurde auch auf einer Skala von 1 bis 5 (1= nicht sinnvoll, 5= sehr sinnvoll) mit einem Wert von 2.6 bewertet. Die User brauchten zwischen 20 und 30 Minuten zum Lösen des Krimis. Der ermittelte SUS Score beträgt 52,667. Wenn ein Wert unter 60 Punkten erreicht wird, sind erhebliche Usability Probleme vorhanden, ab 60-80 Punkten ist die Usability als grenzwertig bis gut zu beurteilen. Daraus lässt sich ziehen, dass die Anwendung MurderMystery bisher noch erhebliche Usability Probleme aufweist. Der gemessene ISONORM-Score beträgt 0,84 Punkte. Um eine gute Nutzerfreundlichkeit zu erreichen wird jedoch ein Mindestwert von 1 vorgegeben. Dazu fehlen noch 1,6 Punkte.

5.4 Auswertung

Aus den Ergebnissen lässt sich entnehmen, dass noch Verbesserungsbedarf bei der Unterhalt-

samkeit der Anwendung besteht, sie jedoch trotzdem schon über eine ausreichende Unterhaltbarkeit verfügt, dadurch, dass der Wert immer noch über der Hälfte der Skala liegt. Hierzu lässt sich sagen, dass als weitere Ergänzung der Applikation eine Aufsplittung der Anwendung in mehrere Schwierigkeitsstufen geplant ist, sodass der User selbst über die Schwierigkeit bestimmen kann. Außerdem würden weitere unterschiedliche Geschichten zu der Unterhaltbarkeit beitragen. Eine weitere Möglichkeit dieses Usability-Problem auszugleichen wäre, die Wiedergabegeschwindigkeit zu personifizieren. Um die Ergebnisse einzugrenzen sei gesagt, dass die Studie zu einem sehr frühen Zeitpunkt der Entwicklung durchgeführt wurde, sodass diese Ergebnisse nicht mehr den aktuellen Stand aufzeigen. Nach dem Zeitpunkt der Durchführung der Studie wurden noch erhebliche Veränderungen und Verbesserungen vorgenommen, sowohl an Regular Expressions als auch generell an der Nutzerführung.

6 Diskussion

6.1 Bisherige Anwendung

In der bisherigen Anwendung ist es möglich, einen recht flüssigen Dialog mit Zeugen, Watson und dem Erzähler zu führen. Durch die Nutzung von Regular Expressions ist jedoch eine Kommunikation mit kleinen Einschränkungen gegeben. Aktuell ist die Anzahl der Fragen, die man einem Zeugen stellen kann auf fünf begrenzt. Zwar sind diese in der benötigten Formulierung recht frei, aber es wird auf gewisse Schlagwörter wie "relationship" gehört und reagiert was die Usability beschränkt. Dadurch, dass der Fokus der Anwendung das Hörverständnis ist, ist die Umsetzung von einer guten Usability problematisch. Um dem User ausreichend Feedback darüber zu geben, an welchem Punkt er sich befindet, kommt es unvermeidbar zu Wiederholungen, was die Usability weiter einschränkt aber für eine sinnvolle Gestaltung des Spiels nötig ist. Des Weiteren ist es möglich, jegliche Räume der Anwendung zu betreten, darin Zeugen zu befragen, Räume auf Hinweise zu durchsuchen sowie diese auch wieder zu verlassen. Als Hilfestellung und zur Sicherstellung der Usability tritt Watson ein, der sowohl noch zu durchsuchende Räume als auch Zusammenfassungen liefert.

6.2 Mögliche Erweiterungen der Anwendung

Nächste Schritte wären, einen Vokabeltrainer einzufügen, sodass bei Watson abgefragte Vokabeln außerhalb des eigentlichen Spiels trainiert und wiederholt werden können. Dies würde dazu beitragen, dass nicht nur das Hörverständnis, sondern auch der Wortschatz des Nutzers erweitert wird. Als weiterer Zusatz sind weitere Geschichten geplant, die gegebenenfalls auch durch die Community der User selbst verfasst werden können, da das Skript so aufgebaut ist, dass die verschiedenen Variablen durch einspeisen von einer Datenbank ganz einfach belegt werden können.

Um den Gaming-Aspekt noch weiterzubringen soll es möglich sein einen "Murder of the day" zu ernennen, der Spieler der es in der kürzesten Zeit oder mit den wenigsten Fragen geschafft hat, den Mörder zu enttarnen. Hierbei wäre dann ein weiteres Ziel eine Online-Community durch eine Website oder gegebenenfalls eine eigene App aufzubauen.

Damit die Nutzerbedürfnisse und Individualität gefördert wird, soll der Nutzer in Zukunft auswählen können, wie lange die Geschichte dauern soll oder den Schwierigkeitsgrad auswählen. Dies kann von Spiel zu Spiel neu gewählt werden. Was zum Beispiel durch prozedural generierten Geschichten, aber auch durch eine große Datenbank an Geschichten umgesetzt werden könnte.

Idealerweise könnten Abläufe wie das Befragen der Zeugen mittels Natural Language Processing Tools oder auch künstlicher Intelligenz flüssiger gestaltet werden, da Interaktion zwischen dem User und der Anwendung der Schlüssel zu einem erfolgreichen Anwendungserlebnis sind.

References

- A. Anderson and T. Lynch. 1988. *Listening: Language Teaching: A Scheme for Teacher Education*. Oxford University Press.
- J. Brooke. 2013. *Sus: a retrospective*. volume 8.
- S. Drobez. 2014. *Der Kriminalroman und sein gesellschaftlicher Stellenwert*.
- W. Elley. 1989. [Vocabulary acquisition from listening to stories. Reading Research Quarterly 24\(2\):174–187. https://doi.org/10.2307/747863](https://doi.org/10.2307/747863).
- K. Figl. 2009. [Isonorm 9241/10 und iso-](https://nfdi.ub.uni-due.de/nfdi4text/urn:nbn:de:hbz:5:1-63862-p0011-7)

metrics: Usability-fragebögen im vergleich.
<https://doi.org/10.1524/9783486598551.143>.

- L. Goldman. 1998. *Childs play myth, mimesis and make-believe*. Berg.
- S. Gramley. 1976. Ein programm für das hörverständnis an der hochschule. AAA: *Arbeiten aus Anglistik und Amerikanistik* .
- T. Kesselring. 1999. *Jean Piaget*. Verlag C.H.Beck.
- R. Miranda. 2019. Interaktive geschichten erleben mit alexa ein comeback. <https://tinyurl.com/v7v7dv6>.
- G. Nold and H. Rossa. 2007. *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)*., Beltz, Weinheim u.a., pages 178–196. <https://nbn-resolving.org/urn:nbn:de:0111-opus-31405>.
- P. Nusser. 2009. *Der Kriminalroman*. J.B. Metzler, Stuttgart. <https://doi.org/10.1007/978-3-476-05236-0>.
- H. Soleimani and M. Akbari. 2013. The effect of storytelling on children's learning english vocabulary: A case in iran. *International Research Journal of Applied and Basic Sciences* Vol, 4(11):4005–4015. www.irjabs.com.
- C. Turner, J. Lewis, and J. Nielsen. 2006. *Determining Usability Test Sample Size*, volume 3, pages 3084–3088.