„OurPuppet“ – Entwicklung einer Mensch-Technik-Interaktion für die Unterstützung informell Pflegender

Todor Dimitrov und Oliver Kramps
Anasoft Technology AG, Bochum

Edwin Naroska, Julia Demmer und Tobias Bolten
Hochschule Niederrhein, Krefeld

Christian Ressel und Stefan Könen
Hochschule Rhein-Waal, Kamp-Linfort

Tim Polzehl und Jan-Niklas Voigt-Antons
Technische Universität Berlin, Berlin

Olaf Matthies und Amir Habibi
Matthies Spielprodukte GmbH & Co. KG, Hamburg

Dominic Heutelbeck und Jana Mertens
Forschungsinstitut für Telekommunikation e.V., Dortmund

Eva-Maria Matip
Deutsches Rotes Kreuz, Bochum

Abstract—In diesem Beitrag wird die technische Umsetzung einer interaktiven Puppe beschrieben, die bei Hochaltrigen und Menschen mit Demenz zum Einsatz kommt. Das Hauptziel ist die Entlastung informell Pflegender, in dem die Puppe beruhigend auf die zu Pflegende Person einwirkt, sie aktiviert und Orientierung im Tagesablauf anbietet. Das System besteht aus der Roboterpuppe, einer zentralen Recheneinheit, einer Backendinfrastruktur und der Smartphone App für die Angehörigen. Die Puppe kann Sprache und Emotionen über die Gesichtsmimik wiedergeben. Außerdem ist sie in der Lage, Menschen im Raum mit dem Blick zu folgen. Rechenintensive Aufgaben wie Sprach- und Emotionserkennung, Kontext-erkennung und -Management und Handlungsplanung finden auf der zentralen Recheneinheit statt. Dort werden aus Rohdaten (Sprache, Gesichtsbilder und Umgebungssensordaten) Kontexte inferiert und der Handlungsplanung bereitgestellt. Diese entscheidet welche vordefiniert Programme ausgeführt werden müssen, um die Puppe zu steuern (z. B. Person ansprechen, an Termine erinnern). Die erkannten Situationen und ausgeführten Aktionen werden im Backend gespeichert und den informell Pflegenden über die App bereitgestellt.

Keywords—Demenz, Roboter, Puppe, Handlungsplanung, Kontexterkennung, Kontextmanagement, Spracherkennung, Emotionserkennung, Dialogführung,

I. EINLEITUNG

Der Einsatz von Robotern in der Pflege wird seit längerer Zeit erprobt. Bekannte Beispiele sind die Robbe Paro und der Spielzeughund Biscuit des Herstellers Hasbro. Beide Roboter werden hauptsächlich eingesetzt, um Menschen mit Demenz (MmD) zu beruhigen und die Konversation mit den Pflegenden Angehörigen (i. F. PFA) und dem Pflegepersonal zu fördern. Die Funktionalität ist beschränkt auf die Erkennung von Berührungen, die Wiedergabe von Lauten und wenigen Bewegungen. Ein deutlich komplexeres System bildet der Roboter Pepper, der im Pflegekontext zur Aktivierung und Sturz Prävention durch körperliche Übungen eingesetzt werden kann. Dieses System kann Dialoge führen und sogar Emotionen erkennen.

Das in diesem Beitrag beschriebene OurPuppet System (i. F.: OPS) vereint die Vorteile beider Robotertypen. Vor der Funktionalität her ähnelt es Pepper ohne auf die Haptik und User Experience der anderen Roboter zu verzichten. Ein wichtiges Unterscheidungsmerkmal ist, dass OurPuppet für die Entlastung der PFA konzipiert wurde und bei kurzer Abwesenheit die zu pflegende Person (i. F. ZPF) unterstützen, beruhigen, aktivieren und ihr Orientierung bieten soll.

II. HINTERGRUND UND ZIELSETZUNG

Das BMBF geförderte Projekt OurPuppet befasst sich mit der Entwicklung und Erprobung einer interaktiven Puppe für den Einsatz im häuslichen Umfeld von Hochaltrigen und Menschen mit Demenz. Durch eine optimierte Mensch-Technik-Interaktion sollen die informell PFA entlastet werden, indem die Puppe beruhigend auf die ZPF einwirkt, sie zu Aktivitäten anregt und Hilfe bei der Tagesstrukturierung anbietet.

Das OPS soll hauptsächlich bei kurzen Abwesenheiten des pflegenden Angehörigen zum Einsatz kommen. Ziel des Vorhabens ist herauszufinden, in wieweit die Puppe einen positiven Einfluss auf die Situation der ZPF nehmen kann. Idealerweise kann der Einsatz dazu führen, dass Stress, Angst oder Unruhe während des Alleinseins von vornherein vermieden werden können. Dies sollte zu einem erhöhten Sicherheitsempfinden des PFA beitragen.

III. METHODEN

Die Umsetzung des OPS Ansatzes wird durch den Einsatz verschiedener Systeme vor Ort der ZPF und in der Cloud

ermöglicht. Deren Architektur und Funktionsweise wird im Folgenden erklärt.

A. Gesamtarchitektur

Die Gesamtlösung besteht aus der Puppe, einer zentralen Recheneinheit (Home Gateway, i. F.: HG), dem Internetserver und einer Smartphone App für die PFA. Die Puppe besitzt Sensoren für die Bestimmung der Lage (liegend, sitzend, wird getragen) und fürs Tracking/Verfolgung der Personen im Raum. Über die Gesichtsmimik kann sie verschiedene Emotionen ausdrücken und ermöglicht so eine natürliche Dialogführung. Unterstützt wird diese durch eine emotional angereicherte Sprachsynthese.

Die Kontexterkennung findet auf dem HG statt. Dort werden Informationen über die Puppe, die Umgebung und die Zielperson inferiert und für die weiteren Komponenten bereitgestellt. So kann u. a. die Handlungsplanung diese Informationen auswerten und geeignete, vordefinierte Programme ausführen, um die Aktionen der Puppe (Person ansprechen, an Termine erinnern, etc.) zu steuern. Die Dialogführung nutzt eine grammatik-basierte Sprach-erkennung, die ebenfalls auf der zentralen Recheneinheit ausgeführt wird.

Um eine möglichst genaue Einschätzung des Gemüts-zustands der ZPF zu erhalten, nutzt das OPS Emotionserkennungsalgorithmen. Zum einen werden aus dem Sprachsignal, welches während der Dialogführung aufgenommen wird, Merkmale extrahiert und die momentane Emotion klassifiziert. Zum anderen erkennt das System „Facial Landmarks“ in den Gesichtsbildern und bestimmt so mithilfe eines neuronalen Netzes die Emotion.

Die PFA sind stets in der Lage sich ein Bild über den aktuellen Zustand der ZPF zu verschaffen. Über die Smartphone-App können sie sich über den Tagesablauf informieren und sowohl Zustandsänderungen als auch von der Puppe ausgeführte Handlungen einsehen. Die Daten werden im Internetserver aufbereitet und über Zugriffsschutz-mechanismen den PFA zur Verfügung gestellt. Auf dem Server laufen ebenfalls Prozesse zur Unterstützung der Kontaktaufnahmefunktion.

Das Zusammenspiel aller Komponenten ist in Abb. 1 dargestellt. Die Puppe baut eine permanente „Socket“-Verbindung zum HG auf, über die sowohl Kontext-informationen (z. B. Lage der Puppe) als auch Rohdaten (z. B. Sprache, Gesichtsbilder, etc.) übertragen werden. Zusätzlich können in beiden Richtungen RPC-Methoden aufgerufen werden, um bspw. die Mimik der Puppe zu steuern.

Das HG hat ebenfalls über das Internet eine permanente Verbindung zum OurPuppet Backend. Darüber werden wichtige Ereignisse protokolliert und dem Nutzerprofil zugeordnet. Auch die Kommunikationsfunktionen werden dadurch ermöglicht. Zusätzlich wird die Backendinfrastruktur eingesetzt, um die Fernwartung und das Monitoring der Systeme bei den ZPF zu ermöglichen. So können Updates

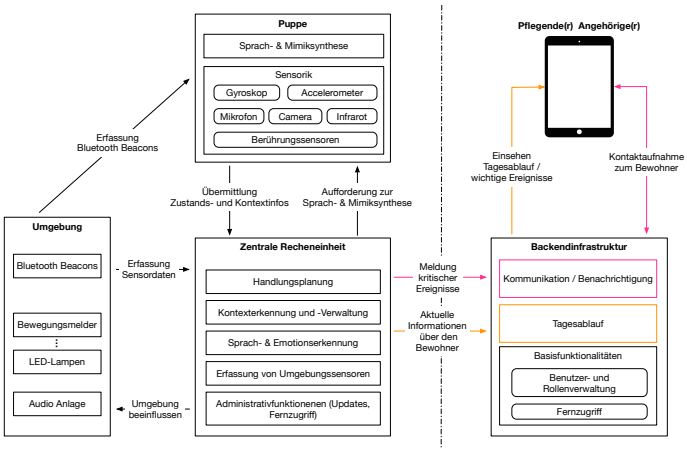

Abb. 1: Gesamtsystem

ausgeliefert werden und sogar bei Bedarf bzw. Wunsch der Nutzer die OurPuppet-Funktionen aus der Ferne ausgeschaltet werden.

B. Puppe

Die Puppe enthält zahlreiche Sensoren und Aktoren. Die Sensoren dienen dazu Informationen über die ZPF zu sammeln und mit Hilfe der Aktoren in Interaktion zu treten. Unter anderem enthält die Puppe:

- Ein im Kopf der Puppe verbautes Mikrofon mit einer ausgeprägten Richtcharakteristik. Mit Hilfe der Fähigkeit der Puppe ihren Kopf zu drehen und Gesichter zu erkennen, richtet sie bei einer Kommunikation den Kopf und damit das Mikrofon auf die ZPF aus. Diese Ausrichtung emuliert nicht nur die natürliche Verhaltensweise von Menschen während der Kommunikation, sondern verbessert aufgrund der Richtcharakteristik des Mikrophons gleichzeitig auch den Signal-Rausch-Abstand.

- Eine Kamera mit einer Fischaugen-Optik um einen möglichst großen Bereich der Umgebung erfassen zu können. Die Kamera ist dabei auf dem Kopf der Puppe (als Brosche/Schleife) angebracht. Die Kamerabilder werden aber vom System nicht aufgezeichnet, sondern nur zur Erkennung von Gesichtern bzw. Analyse der Gesichtszüge genutzt.

- Ein Thermopile-Array (IR-Wärmesensor), welches die Wärmesignatur der Umgebung als ein niedrig aufgelöstes Bild aufnimmt. Hierbei werden lediglich 1024 Pixel aufgelöst, so dass keine Details sichtbar sind. Der IR-Wärmesensor wird dazu genutzt, um Personen und Gesichter verfolgen zu können, ohne eine aufwändige Bildanalyse durchführen zu müssen.

- Beschleunigung-, Lage- und Berührungssensoren werden genutzt, um haptische Interaktionen der ZPF mit der Puppe zu erkennen. Damit soll z. B. erkannt werden, wenn die ZPF sie anfasst oder trägt.

- Mit Hilfe eines Bluetooth-Empfängers kann die Puppe innerhalb der Wohnung verbaute „Bluetooth-Beacons“ identifizieren. Darüber ist es möglich die Puppe grob zu lokalisieren, um so z. B. herauszufinden, ob sie sich im Badezimmer oder in der Küche aufhält.

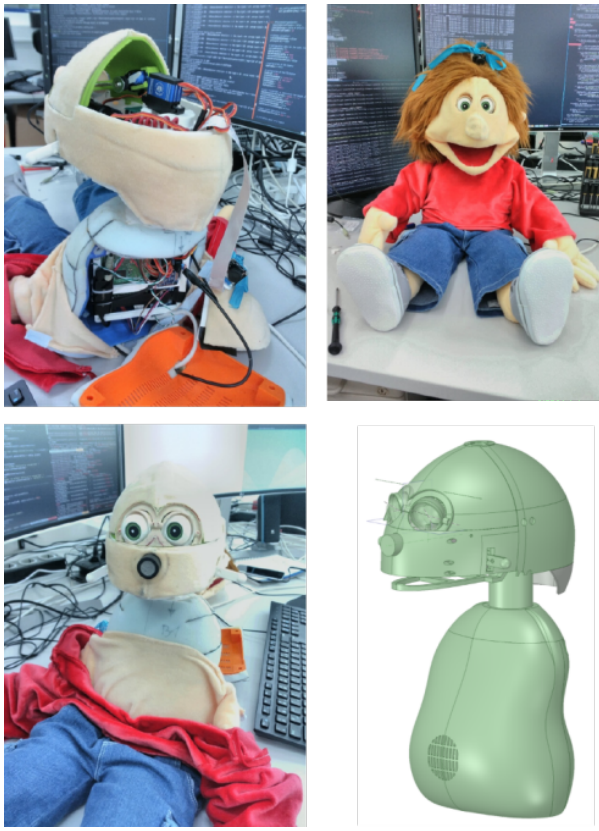

Abb. 2: Aufbau der Puppe

Die für den Betrieb der Puppe erforderliche Energie wird über einen in der Puppe verbauten Akku zur Verfügung gestellt. Damit bleibt die Puppe mobil und kann von der ZPF mitgenommen werden, während er sich durch die Wohnung bewegt. Diese Funktion bzw. Anforderung stellt gleichzeitig auch eine große technische Herausforderung dar. Aufgrund des nur begrenzt zur Verfügung stehenden Bauraums Abb. 2, der Gewichtsbeschränkungen sowie der angestrebten autarken Laufzeit von mindestens einem halben Tag, sind die Energiereserven sehr begrenzt. Daher sind aufwändige und damit energieintensive Berechnungen auf der Puppe nur für einen kurzen Zeitraum möglich. Das Aufladen des Akkus erfolgt über eine fest in der Wohnumgebung verbaute Ladestation.

Die im Puppenkörper verbauten Aktoren dienen im Wesentlichen dazu mit der ZPF zu interagieren:

- Mit Hilfe eines TTS-Systems („text-to-speech“) wird Sprache synthetisiert und über einen im Bauch verbauten Lautsprecher ausgegeben. Zusammen mit der Spracherkennung wird so ein Dialog zwischen der ZPF und Puppe realisiert.

- Ein mechanisch ansteuerbarer Mund ermöglicht die Synthese von Emotionen. Im Detail kann zum einen der Unterkiefer der Puppe sowie die Mundwinkel bewegt werden. Als Ergebnis kann die Puppe Lächeln oder auch „traurig gucken“. Weiterhin wird der Unterkiefer während der Sprachausgabe passend animiert.

- Der Kopf der Puppe kann nach links und rechts geschwenkt werden, damit die Puppe zum einen bei der Sprachinteraktion die ZPF direkt anschauen und zum anderen auch den Raum absuchen kann.

- Die Augen der Puppe sind ebenfalls animiert. Im Detail können die Augenlider bewegt sowie die Augäpfel nach rechts und links geschwenkt werden. Ursprünglich wurden Displays genutzt, um die Augen darzustellen. Allerdings haben Analysen der ersten Prototypen gezeigt, dass mechanische Augen einen deutlich lebendigeren Eindruck erzeugen.

Die Puppe wird kontrolliert von einem Raspberry-Pi-System auf dem Linux läuft. Eine Datenverbindung zum Gateway wird über WLAN aufrecht gehalten, während Bluetooth lediglich zur Erfassung der Bluetooth-Beacons eingesetzt wird.

Eine zentrale Funktion der Puppe ist die Erkennung und Verfolgung von Gesichtern. Dank ausgefeilter Verfahren und Techniken kann dies zurzeit sogar auf kleinen und mobilen System in Echtzeit durchgeführt werden. Allerdings sind die dafür erforderlichen Rechenressourcen und der damit verbundene Energieaufwand nicht unerheblich.

Tabelle 1 vergleicht die Laufzeiten für zwei schnelle Verfahren zur Gesichtserkennung auf einem Raspberry-Pi 3 für zwei verschiedenen Auflösungen des Eingangsbildes. Zwar sind Laufzeiten von bis zu 7 Frames pro Sekunde erzielbar, allerdings gilt dies nur für eine niedrige Auflösung und zudem versagen die Verfahren, wenn die Gesichter nicht frontal zu erkennen sind. Qualitativ bessere Verfahren (z. B. basierend auf Neuronalen Netzen wie YOLO [1]) erfordern leider auch eine erheblich höhere Rechenleistung. Als Ergebnis sind die aktuellen Verfahren nicht geeignet, um eine Erkennung mit nur eingeschränkten Energieressourcen sicherzustellen. Daher wurde ein alternatives Konzept für das Personen- und Gesichts-Tracking entwickelt, welches eine Kombination der Wärme- und Bilddaten verwendet.

Abb. 3 zeigt ein Gesicht einmal als Kamera und einmal als Wärmebild. Deutlich ist zu erkennen, dass das Gesicht relativ viel Wärme abstrahlt und sich daher typischerweise vom Hintergrund deutlich abhebt. Zudem besteht das Wärmebild nur aus 1024 Pixeln, sodass die Übertragung und Verarbeitung der Wärmedaten deutlich effizienter ist als die der normalen Kamera.

Gleichzeitig kann aber über die Wärmesignatur nicht eindeutig erkannt werden, ob es sich um einen Menschen handelt oder um eine andere Wärmequelle. Um hier eine Person von anderen Wärmequellen unterscheiden zu können, wird daher eine Analyse des herkömmlichen Kamerabildes durchgeführt, während einmal identifizierte Personen dann anhand der Wärmesignatur verfolgt werden können. Als Ergebnis kann die Kamera daher die meiste Zeit deaktiviert werden, was zu einer deutlichen Reduktion des Energiebedarfs führt.

Damit die Puppe von der ZPF als Begleiter im Alltag akzeptiert wird, soll sie auf einer emotionalen Ebene mit der ZPF

Tabelle 1: Laufzeit von Haar- und LBP-basierten Gesichtserkennungs-verfahren (Raspberry PI 3)

Bildauflösung in Pixeln	***Laufzeit Haar-Cascades***	***Laufzeit Local Binary Patterns***
640 x 320	approx. 850 ms	approx. 350 ms
320 x 240	approx. 350 ms	approx. 150 ms

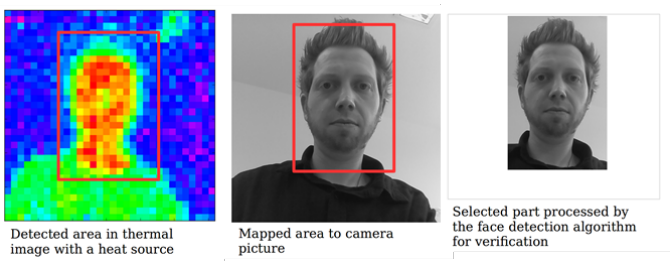

Abb. 3: Gesichtserkennung und Verfolgung

interagieren können. Dazu kann die Puppe Emotionen verbal ausdrücken. Wir verwenden dabei ein TTS-System von der Acapela-Group [9] welches u. a. über qualitativ hoch-wertige deutsche Kinderstimmen verfügt. Neben dem Text können dabei verschiedene „vocal smileys“ generiert werden (Lachen, Weinen, etc.), wodurch die Sprachsynthese erheblich an Lebendigkeit gewinnt.

C. Kontextmanagement

Die Kontextinformationen werden beim OPS in einer probabilistischen RDFS Wissensbasis verwaltet. Dazu werden in der Basisontologie zunächst nur sichere Aussagen modelliert. Erst zur Laufzeit können Aussagen mit Wahr-scheinlichkeiten behaftet werden, z. B. der erkannte emotionale Zustand der ZPF ist zu 70% neutral. Die Inferenz und die Propagierung der Unsicherheit erfolgen mithilfe der „Semi-Naive Datalog“ Evaluierung, wobei die RDF(S) „Entailment Rules“ zur Anwendung kommen. Der Algorithmus ist im OurPuppet Backend in einer Postgres DB mittels PSQL implementiert. Als Ergebnis werden sowohl die neu inferierten RDF „Triples“ als auch die dazugehörigen „Binary Decision Diagrams“ (BDD) für die Berechnung der Wahrschein-lichkeiten zurückgegeben.

Ein Vorteil des gewählten Ansatzes ist, dass bei Wahrscheinlichkeitsänderungen von „Triples“ die Inferenz nicht erneut durchgeführt werden muss. Lediglich die neue Evaluation der betroffenen BDDs ist erforderlich und dies kann sehr effizient erfolgen.

Das Kontextmanagement ist als ein typisches Pub-Sub-System umgesetzt. Die Kontextpublisher sind verantwortlich für die Bereitstellung und Aktualisierung von Kontext-informationen. So existieren im OPS diverse Publisher, welche dediziert Kontexte aus den Umgebungs- und Puppensensoren bestimmen und für die anderen Systemkomponenten, insb. die Handlungsplanung, bereithalten. Als Beispiele können die Bestimmung der Lage der Puppe und die Ortung der Personen in der Wohnung genannt werden, wobei im ersten Fall der Publisher innerhalb der Puppe umgesetzt ist. Die Kontextinformationen können entweder mittels einer SPARQL-Schnittstelle abgefragt werden oder mithilfe eines „Subscribers“ an die interessierten Komponenten weitergeleitet werden. Die Subscriber müssen relevante Kontextänderungen konfigurieren z. B. eine Benachrichtigung senden, wenn die Wahrscheinlichkeit einer bestimmten erkannten Emotion einen Schwellwert übersteigt.

D. Handlungsplanung

Jeder Interaktion zwischen PFA und ZPF wohnt eine zielgerichtete Intention inne. Intentionen können das Beruhigen oder die Anregung bzw. Anleitung einer Tätigkeit oder das Erinnern sein. Beispielhaft seien an dieser Stelle die Anregung zum Trinken, zum Bekleiden, das Erinnern an einen Termin oder das Anleiten während des Essens genannt.

Auch das technische OPS soll, wie ein PFA, unterschiedliche Intentionen bei der Kommunikation und dem Umgang mit der ZPF verfolgen können. Das System soll dabei motivierend und zielgerichtet vorgehen. Diesen Aspekt des Systems realisiert die Handlungsplanung des OPS.

Grundlegendes Ziel der Handlungsplanung ist die sinnvolle Kombination von Einzelaktionen, um ein dem vorliegenden Zustand angemessenes Gesamtziel zu erreichen. Insgesamt soll das System zielgerichtet auf den Zustand der ZPF reagieren und zu einer Entlastung des PFA beitragen. Um dieses Ziel zu erreichen, muss das System sowohl auf direkte Anfragen (Ansprechen) als auch auf auftretende Situationen ohne verbale Äußerungen reagieren können. Eine entsprechende Bewertung der vorliegenden Situation bzw. der Anfrage ist daher unumgänglich, ehe eine Reaktion ausgelöst wird. Das OPS implementiert daher verschiedene Kanäle zur Informationsaufnahme. Für die Kommunikation mit der ZPF sind hier vor allem die Spracherkennung, die Lage-, Bewegungs- und Berührungssensoren der Puppe zu nennen. Zur Bewertung ist eine Bündelung der Informationen dieser unterschiedlichen Kanäle notwendig. Weitere Sensoren im Umfeld der ZPF können detaillierteren Aufschluss über die aktuell zu erfüllenden Erwartungen und Bedürfnisse geben und die akute Zielsetzung somit verfeinern.

Die Planung von Einzelaktionen zu einem Plan, um ein Ziel zu erreichen ist bereits für ungeübte und ungelernte Menschen eine Herausforderung. Dies gilt auch für die Kommunikation mit Menschen mit leichter bis mittelschwerer Demenz. Eine automatisierte Planung von Aktionen und akzeptabler Kommunikation ist daher nicht trivial, auch wenn bereits entsprechende Ansätze in [2], [3], [4] evaluiert wurden.

Deswegen nutzt die Handlungsplanung in ihrer ersten Ausprägung ein in den Möglichkeiten der Kombination von Einzelaktionen eingeschränktes System, bei dem Handlungsstränge, sogenannte Handlungsprogramme, durch Experten vordefiniert wurden. Durch die Kombination von Handlungsprogrammen können vollständige Szenarien abgebildet werden.

Ein Handlungsprogramm umfasst die folgenden Abschnitte: einen Identifikator; eine Liste der benötigten Komponenten des Systems; Regeln die zur Ausführung des Programms erfüllt sein müssen („Preconditions“); eine Beschreibung der Abfolge von Einzelaktionen. Die Beschreibung der Abfolge von Einzelaktionen erfolgt in Form eines gerichteten nicht zyklischen Graphen. In diesem Graphen bilden Knoten auszuführende Einzelaktionen und Kanten Übergänge zu anderen Aktionen, die eine zu erfüllende Bedingung enthalten können.

Beispielhaft ist eine Visualisierung des Ablaufs des Handlungsprogramms „Erinnern“ in Abb. 4 dargestellt. Das Programm unterscheidet zwischen den drei Dringlichkeits-stufen niedrig, mittel und hoch. Das Programm enthält eine entsprechende Verhaltensanweisung für ein positives/negatives

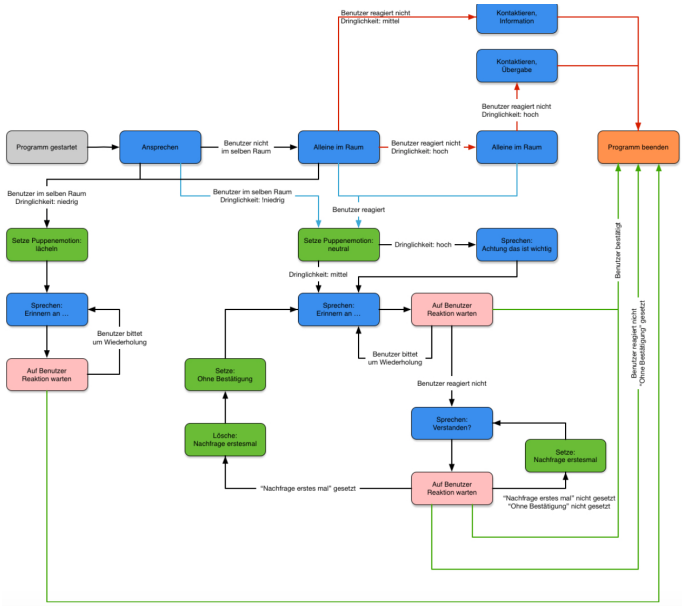

Abb. 4: Ablauf des Handlungsprogramms "Erinnern"

aber auch ein Nicht-Reagieren der ZPF auf den Erinnerungsversuch.

Zur Ausführung kommen Handlungsprogramme in einer „Runtime“ (i. F.: RT), welche den Umfang der möglichen Einzelaktionen bereitstellt. Sie übernimmt ebenso die Kommunikation zwischen dem Handlungsprogramm, dem „Traveler“ (Positionsmarke im Graphen) und dem System. Die Interaktion der RT und den Komponenten außerhalb ist über verschiedene Dienste in der Laufzeitumgebung des Systems realisiert.

Eingaben werden von einer Instanz entgegengenommen, bewertet und an die aktuell aktive RT weitergereicht. Hier ist auch die Verantwortlichkeit, die bekannten Regeln zur Ausführung von Programmen zu bewerten und falls nötig eine neue RT zu erstellen um eine neues Handlungsprogramm auszuführen.

Aus der bisher beschriebenen Architektur ist ersichtlich, dass zu einem beliebigen Zeitpunkt, mehr als eine RT und damit mehr als ein Handlungsprogramm ausgeführt werden könnte. Eine Verwaltung aller ausgeführten Handlungs-programme übernimmt ein „FrameStack“ welcher sicherstellt, dass zu einer Zeit immer nur ein Programm aktiv ist. Nach Beendigung des aktuell aktiven Handlungsprogramms, wird das zeitlich vorangegangene Programm reaktiviert.

Diese dargestellten Strukturen erlauben eine Vielzahl möglicher Szenarien die durch Programme abgedeckt werden können. Eine Erweiterung der Funktionalität ist in einfacher Weise über die Erstellung neuer Handlungsprogramme möglich.

E. Spracherkennung

Die Sprache wird mit dem Richtmikrofon der Puppe aufgenommen und zum HG übertragen. Dort erfolgt die Spracherkennung mithilfe eines grammatik-basierten Erkenners, wobei die eigentliche Audiotranskription mithilfe eines „Recurrent Neural Networks“ erfolgt. Die Terminierung der Aussagen wird durch einen „Keywordspotter“ und einen „Voice-Activity-Detector“ (VAD) unterstützt. So muss die Puppe stets mit ihrem Namen („Elisa“) angesprochen werden. Um sicherzustellen, dass Hintergrundgeräusche (z. B. Radio) nicht dazu führen, dass die Erkennung permanent läuft, wird jede angefangene Aussage nach einigen Sekunden automatisch terminiert, falls der VAD keine Silence detektiert.

Die erkannten Phrasen/Sätze werden der Handlungsplanung übergeben und dienen dort als Trigger bzw. Kontext-informationen für die Ausführung der Programme. Eine natürliche Dialogführung wird ermöglicht, in dem bei erwarteten Antworten für eine kurze Zeit auf den Keyword-spotter verzichtet und die Sprache direkt zum Grammatik-erkenner weitergeleitet wird.

F. Emotionserkennung per Sprache

Ist ein Sprecher beispielsweise ängstlich, so ist zu erwarten, dass sich seine Sprechweise auditiv deutlich von der eines freudigen oder gelangweilten Sprechers unterscheiden lässt. Im OPS werden die emotionalen Zustände wie Ärger, Freude, oder Überraschung anhand der spezifischen vokalen Ausdrucksweise im Umgang mit bspw. der Tonhöhe, Sprechgeschwindigkeit, Sprechrhythmus und Stimmklang oder Lautheit der Sprecher erkannt.

Das wohl am häufigsten verwandte akustische Merkmal ist die Grundfrequenz, welche im OPS mittels einer Form der Autokorrelationsfunktion, kurz AKF, extrahiert und auf wahrnehmungsbezogene Einheiten wie Halbtonschritte, Mel- oder ERB-Skala dargestellt wird. Hier wählt das OPS eine Normierung anhand der Mittelwerte der globalen Äußerung in der Annahme einer mittleren Ausprägung bei neutral gesprochener Sprache. Die so ermittelten Kurven sind oft verrauscht und mathematisch schwer zu beschreiben. Dem entgegenwirkend werden verrauschte Signale im OPS durch einen Filter geglättet. Häufig verwandte Merkmalsgrößen dieser Parametrisierung sind statistische Kenngrößen wie Extrema, Momente erster und zweiter Ordnung sowie Deltawerte. Weiterhin werden Pausen und Dauern im Sprachsignal aus dem Verlauf der Signalenergie geschätzt. Zudem bestehen eine Reihe komplexer und umstrittener psychoakustischer Zusammenhänge zwischen der Empfindung der Lautheit eines Tones, der Frequenz und der Dauer desselben. Für das Verhältnis spektraler Bänder kann auf „Mel-Frequency-Cepstral-Coefficients“, kurz MFCC, zurück-gegriffen werden. Häufig wird auch kategorial die Energie unterhalb 250Hz beziehungsweise 650Hz betrachtet. Die harmonische Ausprägung kann mittels der „Harmonic-To-Noise Ratio“, kurz HNR, bestimmt werden.

Aus der Vielzahl an klassifizierenden Systemen haben sich für die Emotionsforschung Tendenzen zu bestimmten Strategien ergeben. Die meisten Systeme benötigen dabei eine Lernphase, in der Parameter gelernt oder optimiert werden, bevor diese in der späteren Testphase mit neuen Mustern verglichen werden können. Aufgrund sehr guter und speziell bei kleinen Datenmengen gegebener Robustheit arbeiten wir mit sogenannten Support-Vektor-Maschinen (SVM) für die Modellierung und Vorhersage der Emotion des Sprechers.

Für das OPS ergab sich eine Akkuratesse von im Mittel 79,63% bei einer Ratewahrscheinlichkeit von 12,5% über eine Test-Evaluierung anhand eines Standardkorpus mit 7 Emotionsklassen: Ärger (Wut), Langeweile, Ekel, Angst, Freude, Trauer, Neutral. Anhand dieses Korpus können Ergebnisse verglichen und als sehr vielversprechend bewertet werden. Während die Ergebnisse für die Klasse „Ärger" in moderaten bis guten Bereich liegen, sind die Ergebnisse für Angst etwas schlechter, und für die Emotion Freude als mittelmäßig einzustufen. Der Grund dafür ist das linguistisches Phänomen, dass die beiden Emotionen Ärger und Freude oft aufgrund einer ähnlich ausgeprägten hohen stimmlichen Aktivierung verwechselt werden.

Zukünftige Herausforderungen gibt es viele. Wird bspw. nicht mit Sprechrichtung hin zur Puppe gesprochen, wird sehr leise gesprochen, oder bewegen sich die Sprecher im Raum, muss damit gerechnet werden, dass die variierende Signalstärke die Erkennung sehr schwierig gestalten wird. Trotzdem ist gerade dies oftmals ein emotionales Zeichen, bspw. wimmern, weinen, flüstern. Letztlich stören physische Hindernisse oder akustische Verdeckungen, bspw. die Bettdecke zwischen Sprecher und Mikrofon oder das verdeckte Sprechen aus dem Kopfkissen heraus die Klassifikation ebenso wie ein Durcheinandersprechen mehrerer Personen.

G. Emotionserkennung per Gesichtsbilder

Das OPS sollte in der Lage sein, Emotionen (Wut, Ekel, Furcht, Glück, Trauer, Überraschung und Neutral) aus frontal gerichteten Gesichtern zu erkennen. Als Features wurden 68 Orientierungspunkte im Gesicht („facial landmarks") mithilfe von OpenCV und Dlib extrahiert. Es wurden sowohl SVM als auch neuronale Netze für die Klassifikation eingesetzt, welche mit insgesamt 8939 Testbildern aus vorhandenen Datenbanken (z. B. [7], [8]) trainiert wurden. Die Inputbilder wurden in 80% Trainingsdaten und 20% Testdaten unterteilt.

Die Emotionserkennung mithilfe der SVM basiert auf dem Algorithmus wie in [5] beschrieben und erreicht eine korrekte Klassifikation in 76% der Fälle.

Den neuronalen Netzen wurden als Input ausschließlich die Orientierungspunkte übergeben. Es wurden FFN („Feed Forward Networks") verschiedener Größen getestet, wobei die höchst erzielte Erkennungsrate bei 84% lag. Im nächsten Schritt werden SIFT-Merkmale aus den Umgebungen der Orientierungspunkte als zusätzliche Input-Features verwendet, wie in [6] beschrieben. Die Ergebnisse liegen zum Zeitpunkt noch nicht vor.

H. Datenschutz

Der Datenschutz steht grundsätzlich im Spannungsfeld zwischen der Sicherheit der Daten, Benutzer-freundlichkeit und Performance der Anwendung. Im Falle der Verwaltung von personenbezogenen Daten (pbD) aus dem medizinischen Umfeld sind selbige besonders schutzwürdig. Daher ist das Sicherheitskonzept bzgl. Zugriffsrechten von z.B. PFA darauf ausgelegt die Verfügbarkeit von pbD sowohl auf der Ebene des Zugriffs durch Benutzer als auch auf der Systemebene zu minimieren. Kernaspekt ist die verteilte Datenhaltung, sodass die pbD, wie Sensor- oder Audiodaten, nicht zentral aggregiert und wenn möglich pseudonymisiert werden sollten. Das führt im Vergleich zu zentralisierten Datenbasen zu erhöhten Latenzen, Kommunikationsaufwand, Protokollkomplexität und ggf. zeitweiser nicht Verfügbarkeit der Daten einzelner Nutzer, stellt aber dafür sicher, dass es sehr schwierig bis unmöglich ist für einen zentralen Akteur ohne Zustimmung Zugriff auf pbD zu erhalten.

IV. DISKUSSION UND AUSBLICK

OurPuppet ist ein vielversprechender Ansatz sowohl die Betreuten als auch die Betreuenden im Alltag zu unterstützen. Gerade die haptische und puppenartige Anmutung spricht – nicht alle, aber eine Vielzahl – der Anwender besonders an und ermöglicht es so eine besondere Beziehung zum Betreuten aufbauen zu können. Dazu sind eine Vielzahl von Technologien in das System integriert worden, die letztlich eine möglichst „lebensechte" Interaktion mit dem Nutzer ermöglichen sollen. Gerade diese Mischung von unterschiedlichsten Technologien zu einem Gesamtsystem macht den innovativen Charakter des Projekts aus.

OurPuppet ist ein komplexes technisches System, bei dem eine Vielzahl von Soft- und Hardware-Komponenten miteinander interagieren müssen, um das gesteckte Ziel – die Beruhigung und Aktivierung des Betreuten sowie die Unterstützung des Betreuenden – zu erzielen. In der aktuellen Phase des Projekts werden daher die Komponenten aufeinander abgestimmt und optimiert, um ihre volle Leistungsfähigkeit entfalten zu können. Das geschieht zum einen mit entsprechenden Labortests, wichtiger sind aber die zurzeit anstehenden Nutzertests, bei denen die Puppe bei mehreren Anwendern in einem realen Setting eingesetzt werden wird. Hierbei soll zum einen die Wirkung der Puppe aber auch die Funktionsfähigkeit des Systems untersucht und optimiert werden.

Nach einer ersten Testphase mit den Anwendern, werden die Tests erweitert, um mehr Daten zur Bestimmung der Wirkung der Puppe zu sammeln. Dabei werden die Puppe über einen Zeitraum von mehreren Monaten bei Anwendern in realen Wohnumgebungen getestet.

LITERATUR

[1] REDMON, J., DIVVALA, S., GIRSHICK, R., FARHADI, A.: You Only Look Once: Unified, Real-Time Object Detection, 2015

[2] BEHNKE, Gregor, et al. A Unified Knowledge Base for Companion-Systems–A Case Study in Mixed-Initiative Planning. In: Proceedings of the International Symposium on Companion Technology. 2015.

[3] BERCHER, Pascal, et al. User-Centered Planning. In: Companion Technology. Springer, Cham, 2017. S. 79-100.

[4] NOTHDURFT, Florian, et al. The interplay of user-centered dialog systems and AI planning. In: Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue. 2015. S. 344-353.

[5] VAN GENT, Paul: Emotion Recognition Using Facial Landmarks, Python, DLib and OpenCV, 2016

[6] EL-DIN, YS, MOUSTAFA, MN., MAHDI, H.: Landmarks-SIFT face representation for gender classification, 2013

[7] EBNER, N. C., RIEDIGER, M., LINDENBERGER, U.: FACES - A database of facial expressions in young, middle-aged, and older women and men: Development and validation, 2010

[8] AIFANTI, N., PAPACHRISTOU, C., DEPOPULOS, A.: The MUG Facial Expression Database, 2010

[9] https://www.acapela-group.de