KOMMUNIKATIVE KOMPONENTEN SOZIALER INTELLIGENZ VON KÜNSTLICHEN KOOPERATIVEN SPIELENDEN

Casey C. Bennett[1,2], *Benjamin Weiss*[3], *Jaeyoung Suh*[1], *Eunseo Yoon*[1], *Jihong Jeong*[1], *Sungmin Yang*[1], *Yejin Chae*[1]

[1]*Department of Data Science, Hanyang University, Seoul Korea,* [2]*College of Computing and Digital Media, DePaul University, Chicago, USA,* [3]*Quality and Usability Lab, Technische Universität Berlin, Germany*
cabennet@hanyang.ac.kr, benjamin.weiss@tu-berlin.de

Kurzfassung: In diesem Beitrag wird ein Forschungsprogramm vorgestellt, dessen Ziel es ist, empirisch kommunikative Systemkomponenten und ihre Ausprägungen zu identifizieren, die maßgeblich für die Wahrnehmung und Zuschreibung von Intelligenz und Intentionalität bei sprachbasierten Systemen zuständig sind. Der Ansatz selbst ist datengetrieben und iterativ: Für die Domäne der kooperativen Spiele wird dazu eine Serie von Studien mit „Don't starve together" durchgeführt, um einen virtuellen Agenten zu entwickeln, der als Avatar nicht nur selbständig agiert, sondern auch mit den menschlichen Mitspielenden sprachlich kommuniziert. Neben dem aktuellen System und der bereits erhaltenen Ergebnisse werden die weiteren Forschungsfragen und geplanten Schritte vorgestellt.

1 Einleitung

Sprachbasierte Interaktionssysteme haben inzwischen einen Reifegrad erlangt, der den Einzug in das tägliche Leben als Smart Speaker, zur persönlichen Assistenz oder als Kundenservice-Bot ermöglicht. Eine zentrale wissenschaftliche Aufgabe der vergangenen Jahre war die Identifikation und Evaluation von Eigenschaften und Verhaltensweisen, die solche Systeme menschlicher und natürlicher wirken lassen, mit der Erwartung, dass dies eine Voraussetzung, oder zumindest ein bedeutenden Aspekt für positive User Experience ist [1, 2].

Allerdings müssen etwa das Erleben von Sozialer Präsenz [3] oder die Zuschreibung von Intelligenz nicht zwangsläufig mit dem Grad an Menschlichkeit oder Natürlichkeit korrelieren [4]. Statt dessen kann für einen kooperativen und erfolgreichen Dialog mit interaktiven Systemen die Zuschreibung von Intentionalität bedeutsam sein [5, 6], also das Verhalten von Systemen begründet und zielorientiert ist, wie es auch die häufige Verwendung des Begriffes *Agent* impliziert. Bei *Intentionalität* handelt es sich jedoch um ein wenig beforschtes Konzept in den Communities zur Sprachinteraktion und Human-Computer- bzw. Human-Robot-Interaction (HCI, HRI). Sie und die verwandte *Handlungsfähigkeit*, also die Fähigkeit zur Intentionalität, werden bspw. in einer Metaanalyse zu zentralen Begriffen in der Forschung zu intelligenten Nutzerschnittstellen gar nicht berichtet [7].

In diesem Beitrag wird deshalb ein Forschungsprogramm vorgestellt, dessen Ziel es ist, für kommunikative Agenten, virtuelle Assistenten und Roboter, solche technischen Komponenten zu identifizieren, die eine Zuschreibung von Intentionalität und Handlungsfähigkeit beeinflussen.

Abbildung 1 – Screenshot des Spieles mit zwei menschlichen Mitspielern und deren Kommunikation über Zoom

2 Forschungsstand

Für dieses Forschungsprogramm wurde die Domäne der kooperativen Spiele ausgewählt, deren Erfolgsorientierung eine gute Grundlage bietet, um einen Mehrwert künstlicher Intelligenz (KI) zu messen [7], sowohl hinsichtlich des Spielergebnisses, aber auch bezogen auf einen implizit wohl-definierten Kontext für subjektive Nutzerurteile. Dieser hier vorliegende Kontext bildet einen begrenzenden und auch manipulierbaren Rahmen, in dem kommunikatives und nicht-kommunikatives *koordiniertes* Handeln stattfindet [8] und erlaubt, die kontextuelle Angemessenheit als eine der zentralen Fähigkeit interaktiver KI zu bewerten [9]. Zudem bietet ein Spiel ausreichend Ereignisse, auf die sprachlich Bezug genommen werden kann und die kooperative Situation verhindert grundsätzlich einige Arten komplexer Phänomene, wie etwa Täuschungen.

Fließende Interaktion ist kulturabhängig und basiert sowohl auf angemessenen verbalen wie nonverbalen Signalen, also bezogen auf Raum, Zeit, Frequenz [10, 11] und natürlich auf einander [12, 13]. Zudem sollte soziales Handeln eine Balance zwischen vorhersehbaren und unvorhersehbaren Verhalten darstellen, um einerseits sozialen Erwartungen an eine erfolgreiche Kommunikation zu entsprechen [14, 10], andererseits Informationsgehalt zu generieren, der für einen selbständigen Akteur bzw. Intentionalität und für den Aufbau eines entsprechenden Mentalen Modells notwendig ist [15].

In diesem Beitrag soll das verbale Verhalten im Mittelpunkt stehen, da auch die bisher durchgeführten und gerade vorbereiteten Studien ausschließlich sprachliche Komponenten zum Forschungsgegenstand haben.

3 Systemumgebung

Als Umgebung für die kooperative Interaktion wurde sich für das Spiel „Don't starve together"[1] entschieden. Dabei handelt es sich um ein Überlebensspiel, wo Ressourcen gesammelt, Werkzeuge erstellt und Tiere/Monster bekämpft werden müssen (Abbildung 1). Mit neuen Werkzeugen können neue Arten von Ressourcen erschlossen werden. Zudem wird durch einen Tag-Nacht-Zyklus Zeitdruck aufgebaut, bis zum Sonnenuntergang einen Schutz (bspw. Lagerfeuer) gesammelt und aufgebaut zu haben.

Von den verschiedenen Szenarien wurde das soziale ausgewählt. Hier kann durch gemeinsames Handeln und Austausch von Besitz das Überleben wesentlich erleichtert werden. Für

[1] https://www.klei.com/games/dont-starve-together

Abbildung 2 – Screenshot des Spieles. Eine Mitspielerin ist durch einem Avatar repräsentiert

unsere Testreihen wurde der Mehrspielmodus auf zwei reduziert, um einem einzigen menschlichen Partner mit dem Agenten interagieren zu lassen. Dazu wurde eine Spielumgebung so konfiguriert, dass ausschließlich eingeladene Nutzer mitspielen können. Zusätzliche Erweiterungen und Konfigurationen werden über spezielle Spiel-Mods (Modifikationen) bereitgestellt oder selber in LUA implementiert. Ein solcher Mod wurde erstellt, um für die Sprachinteraktion Umgebungs- und Spielstanddaten zu erfassen (Logdaten des aktuellen Standes etwa alle 30 Sekunden), bspw. Spielendenzustände (wie etwa Hunger, Lebensenergie), Inventar oder wahrnehmbare Bedrohungen, und um eine nicht vorhandene Pause-Funktion zu erhalten, die für den Start und Abschluss eines Experimentdurchlaufes notwendig sind.. Außerdem wurde auf diese Weise eine feste und vorteilhafte Startposition für die Spielenden eingestellt, die Beleuchtung und nahe Ressourcen umfasst, sowie das untere Limit für die Lebensenergie auf 10% festgelegt. Dieser Mod soll insbesondere für unerfahrene Spielenden die störungsfreie (also todesfreie) Interaktion mit dem Agenten fördern, ohne dass dies den Teilnehmern mitgeteilt wird. Dieser Mod dient auch als Schnittstelle für weitere automatische Steuerungen des Spielfeldcharakters des Agenten (s.u.).

Um die sprachliche Kommunikation kontrollieren zu können, wurde parallel zum interaktiven Spiel über Zoom eine Verbindung aufgebaut, in der über den Zoom-Audiokanal miteinander gesprochen werden kann und ein virtueller Avatar des Agenten (zusätzliche zum Spielcharakter auf dem Spielfeld) über loomielive[2] in den Zoom-Videokanal sichtbar ist (Abbildung 2). Während in der Entwicklungsphase einige Funktionen von geschulten Mitgliedern des wissenschaftlichen Teams übernommen wurden, ist das technische Ziel einen Agenten zu entwickeln, der autonom und zielorientiert . . .

1. auf dem Spielfeld agiert und interagiert,
2. mit dem menschlichen Mitspieler eventbasiert verbal kommuniziert,
3. über den visuellen Zoom-Kanal angemessen nonverbal kommuniziert,
4. und dessen Spielverhalten und Kommunikation kongruent sind.

Selbst bei Reduzierung auf die rein kommunikativen Elemente einer sozialen KI bedarf es an technischen Komponenten für das gewünschte Verhalten mindestens: Automatische Spracherkennung (ASR) und (audio-visuelle) Sprachsynthese (TTS), ein Handlungsplaner, Reaktionsfähigkeiten auf Umgebungs- und Mitspieleraktionen, sowie ein Dialogmanager. Im Folgenden werden die durchgeführten und geplanten Studien und damit einhergehenden technischen

[2]https://www.loomielive.com/

Entwicklungen vorgestellt. Über einen früheren Stand des Projektes wurde in Suh et al. [16] berichtet.

4 Methode

Grundsätzlich sollen die Interaktionen zwischen menschlichen Mitspielenden und dem künstlichen Agenten über Zoom und dem Spielfeld in einem kooperativen Rahmen stattfinden. Die einzelnen Sitzungen wurden auf etwa 30 Minuten festgelegt, nach deren erreichen das Spiel vom Versuchsleiter beendet werden soll, also nicht einfach unterbrochen. Das serverbasierte Spiel selbst und die aktuelle Corona-Situation bilden einen sinnvollen Rahmen, um die Nutzung von Zoom zu begründen. Alle Sitzungen wurden und werden mit OBS Studio[3] aufgezeichnet.

Es werden drei Studien vorgestellt: Die erste diente der Datensammlung menschlichen Verhaltens, da für die Entwicklung der Sozialen KI kein Design- sondern ein datenbasierter Ansatz gewählt wurde. Die zweite Studie testete den visuellen Zoom-Avatar und das automatische Dialogsystem, während die dritte Untersuchung ein Hypothesen-orientiertes Experiment ist, dass gerade abgeschlossen wird.

4.1 Studie 1 – menschliche Vergleichsdaten

Sechs Mitglieder des Forschungsteams spielten nach dem Test des technischen Aufbaues miteinander in Englisch und Koreanisch mehrere Sitzungen von etwa 30 min. Dauer. Diese Daten wurden von zwei Paaren von Studenten (koreanisch und englisch) annotiert (Cohen's $\kappa = 0.72$) und in gemeinsamen Arbeitssitzungen konsolidiert, um Kommunikationsverhalten zu erfassen und zu systematisieren. Dabei wurden Spielkontexte identifiziert, die Kommunikationen verursachten, sowie Äußerungen des Mitspielers. Unterschiede zwischen den koreanischen und englischen Sitzungen konnten nicht festgestellt werden. Als Ergebnis wurde eine einfache Taxonomie zu kontextabhängigen Äußerungen erstellt (Abbildung 3), die so implementiert wurde, dass die Äußerungen von den Logdaten des LUA-Spielemods aktiviert und per TTS realisiert werden können. Eine parallele Implementierung lauscht keyword- bzw. patternbasiert auf Nutzereingaben, um als eine Art Chatbot, beispielhafte Antworten zu erzeugen. Die häufigsten Kategorien beziehen sich bspw. auf „Wohin“, „Monster ist . . . “, „Folge mir“. Beide Module halten einige wenige (meist 3–5) alternative Äußerungen bereit, aus denen zufällig gewählt wird.

Diese Taxonomie bildet auch die Grundlage für ein Modul, dass das nicht-sprachliche Spielverhalten kontrollieren soll (siehe Abschnitt 5).

4.2 Studie 2 – Mensch-Maschine-Interaktion zur Validierung

Die neuen Sprachfähigkeiten des Agenten – Reaktion auf die Spielsituation – wurde in einer zweiten Studie mit acht, wiederum eingeweihten, Teilnehmern getestet. Hierzu wurde der Zoom-Loomie-Avatar angebunden, der sich rudimentär Lippensynchron zum Eingabe-Audio auch visuell bewegt. Das eigentliche Spielverhalten wurde jedoch von einem Mitarbeiter gesteuert. Eine Reaktion auf Äußerungen der Mitspielenden war noch nicht fertiggestellt. Abschließend wurden strukturiert qualitative Rückmeldungen erhoben. Diese umfassten Aspekte wie etwa wann (nicht) Sprache und Verhalten des Agenten zusammenpassten, was nervig oder positiv am Verhalten des Agenten war oder welche technischen Probleme auffielen.

Ergebnis dieser Studie war neben kleineren Überarbeitungen des Sprachverhaltens eine Liste häufiger Kommentare, mit deren Hilfe noch fehlende Komponenten Sozialer KI identifiziert werden konnten. Diese Komponenten umfassen

[3] https://obsproject.com/

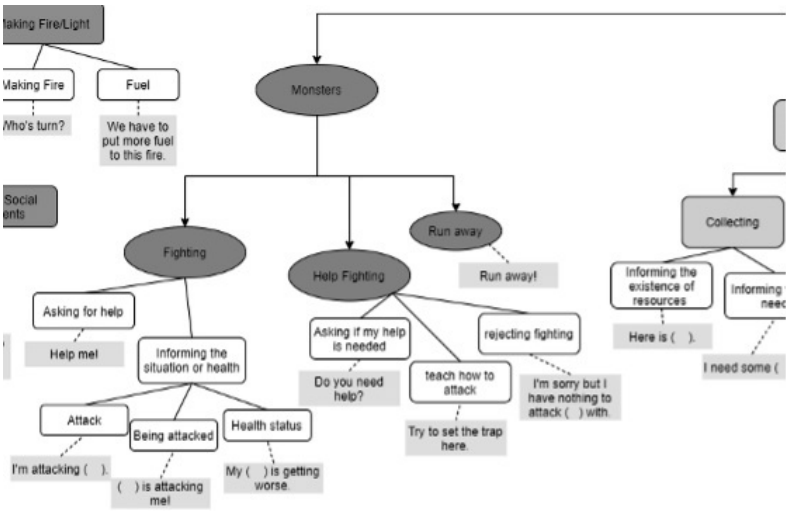

Abbildung 3 – Einfache kontextabhängige Kommunikations-Taxonomie (Teilbaum „Monster“)

1. die bereits angefangene Reaktion auf Äußerungen von Mitspielenden per ASR,
2. die sozial adäquate Verhinderung von Wiederholungen,
3. fehlendes Kontextwissen (explizit geteilte Objekte, Zustand der Mitspielenden),
4. sowie das fehlende Planungsverhalten des Agenten.

Die Komponenten 1. und 3. sind rein technische Natur und konnten schnell ergänzt bzw. mit Anbindung einer API fertiggestellt werden. Insgesamt wurden zu diesem Zeitpunkt 46 verschiedene Kategorien von sprachbezogenen Aktionen definiert.

Für das fehlende Planungsverhalten wird derzeit ein technisches Modul entwickelt und getestet, dass auf Basis der Logdaten und späterer Zustände bzw. Agentenäußerungen für die nahe Zukunft schätzt (siehe Abschnitt 5).

Die 2. Komponente erinnert stark an den Effekt des *Social Inhibition of Return* (IOR), der für menschliche sensorische Funktionen wie etwa dem Sehen bekannt ist. Das nicht nur für Menschen bekannte Effizienz-Prinzip unterdrückt Wiederholungen in zu kurzen Zeiträumen (etwa 1–2 Sekunden) und kann beim Fehlen auf kognitive Störungen hinweisen oder in der HRI zu negativen Nutzererleben führen, etwa dem Eindruck roboterhaften Sprechens oder unmotivierten Unterbrechungen [17]. Deshalb wurde IOR für ein 3 Sekunden-Fenster implementiert, indem weitere erzeugte Äußerungen aus derselben Oberkategorien der Taxonomie nicht abgespielt wurden.

Als letzte Vorbereitung für das Folgeexperiment wurden dem Sprachverhalten Prioritäten zugewiesen (Cohen's $\kappa = 0.61$). So konnten drei Stufen von Sprachaktivität („Geschwätzigkeit“) realisiert werden. Von den 46 Kategorien wurden 14 bzw. 16 und 16 Kategorien den Prioritäten 1–3 zugeordnet. Diese drei Stufen sind:

1. Priorität 1
   - Antworten auf direkte Fragen der Mitspielenden
   - Reaktionen im Einklang mit der Sozialen Norm (bspw. „Hallo“)
2. Priorität 2
   - Kampf-bezogene Inhalte
   - andere Existenz-bezogene Inhalte (Sterben, Verhungern)
3. Priorität 3

- alle weiteren situationsbezogenen Kommentare
- alle Kategorien, die bislang nicht zugeordnet wurden

4.3 Studie 3 – Testen verschiedener Experimentalbedingungen

In dieser ersten Studie mit naiven Teilnehmern werden folgende Bedingungen getestet: IOR ein- oder ausgeschaltet, sowie die drei Prioritäten von Sprachaktivität miteinander verglichen. Für dieses derzeit noch laufende Experiment sind bislang Daten für 32 Teilnehmenden vorhanden, jeweils 16 für die Kontrollbedingung und 16 mit aktivierter IOR. Dabei sind englische und koreanische Sitzungen jeweils zu 50% vertreten. Die Sprachaktivität ist für diese Bedingungen noch als „hoch“ eingestellt.

Mit einem Wizard-of-Oz-Ansatz wird das Spielverhalten des Agenten ohne Wissen der Teilnehmer kontrolliert, während das Sprachverhalten komplett automatisch stattfindet. Zur quantitativen Evaluierung wird der Godspeed Fragebogen [18] und das Networked Minds Instrument [19] zur Erfassung sozialer Präsenz genutzt.

Unsere Hypothese war, dass IOR die Bewertung des Agenten auf beiden Instrumenten verbessern würde. Die ersten Analyse der Fragebögen ergeben jedoch ein anderes Bild. So zeigt sich für die Koreanischen Sitzungen kein Effekt zwischen der Referenz und IOR (Soziale Präsenz: .80 vs. .85, Godspeed: 3.29 vs. 3.24). Für die englischen Sitzungen zeigt sich allerdings der Effekt, dass IOR in weniger Präsenz, aber mehr Sympathie resultiert (Soziale Präsenz: .91 vs. .37, Godspeed: 3.06 vs. 3.51).

Während ein sprachlicher bzw. kultureller Effekt durchaus zu erwarten ist [10, 20], sollte IOR als neuronales Prinzip eigentlich weniger betroffen sein als bspw. Sprachaktivität oder Höflichkeitsverhalten. Mit Abschluss dieses Experimentes wird eine detaillierte Analyse, auch für die bislang nicht betrachten Bedingungen vorgelegt.

5 Laufende Arbeiten und Ausblick

Das eigentliche Spielverhalten ist derzeit noch nicht autonom. Dies wird auf Basis der Taxonomie mit dem FATIMA-Toolkit[4] implementiert. Dazu gehört neben zielorientierten Aktionen auch kontinuierliches nonverbales Raumverhalten, wie etwa das Folgen des Mitspielers oder Umgehen von Hindernissen.

Zudem wird ein datengetriebenes Planungsmodul entwickelt, dass auf Basis von 15 Zeilen Logdaten, also 7-8 Sekunden, aus der zweiten Studie Agentenäußerungen mit einem flachen CNN-LSTM Netzwerk binär für jede Oberkategorie schätzt (1 1D CNN + 1 LSTM 30 units + dense). Aktuell wurden zum Training ausschließlich isolierte Äußerungen ausgewählt, die im betrachteten Zeitfenster vor der Äußerung keine weiteren Sprachaktionen aufwiesen und ausschließlich von Spielsituationen, aber nicht Mitspieleräußerungen, initiiert wurden. Insgesamt wurden 156 solcher Datenpunkte gefunden. 20% der Daten wurden als Testset genutzt. Für diese ergaben sich gemischte Ergebnisse (Accuracy) für die folgenden Oberkategorien: Nacht (.87), Ressourcen (.59), Monster (.85), Erbauen (.88), Soziale Interaktion (.50), externe Events (.84). Insbesondere Soziale Interaktion und Ressourcen konnten nicht vorhergesagt werden. Hier bietet sich an, statt der Vorhersage automatisch generierter Äußerungen eher menschliche Kommunikation zu analysieren und als Strategie, bspw. in Form einer Agenda, zu formalisieren. Da derzeit keine ca. 30–40 Spielstandinformationen betrachtet werden, ist eine Vorhersage, bspw. in den nächsten Sekunden Nahrung zu sammeln, deutlich herausfordernder als eine allgemeine Absichtserklärung für den Fall, dass Nahrung aufgefunden wird.

[4] https://fatima-toolkit.eu/

Auf ein weiteres Modul, dass sich gerade in der Entwicklung befindet, eine visuellen Emotionserkennung, soll hier nicht weiter eingegangen werden. Ziel ist es, diese soziale Kontextinformation für Sprachkommunikation, aber auch zur impliziten Evaluierung des Agenten zu verwenden.

6 Fazit

Ein rudimentärer Agent für kooperative Spiele wurde vorgestellt, der mit Spielstand- und multimodalen Kontextinformationen zu einer autonomen sozialen AI entwickelt werden soll. Zentral bei dieser Entwicklung ist die Ausrichtung auf Intentionalität und Handlungsfähigkeit, die auch von menschlichen Mitspielenden zugeschrieben werden soll. Deshalb bietet sich die spielerische Umgebung mit individuellen Handlungsmöglichkeiten, Kontextinformationen und verbalen wie non- und extra-verbalen Signalen an, während typische aufgabenorientierte Serviceanwendungen in aller Regel nicht die dafür notwendige soziale Rolle und Aufgaben bereitstellen. Die bislang identifizierten Komponenten, *Social Inhibition of Return*, Planungstransparenz und die Nutzung von weiteren Kontextinformationen stellen natürlich nur einen ersten Teil zu erwarteten Komponenten dar. Mit der Erweiterung des Sprachverstehens auf den aktuellen Stand der Technik wird erwartet, weitere, bspw. sprachlich-soziale, Komponenten empirisch identifizieren zu können.

7 Danksagung

Wir möchten uns bei Cheda Stanojevic, Sawyer Collins (Indiana University) und Sungmin Yang (Hanyang University) für ihre Unterstützung bedanken. This research was supported through funding by a grant from the National Research Foundation of Korea (NRF) (Grant number: 2021R1G1A1003801).

Literatur

[1] LESTER, J. C., S. A. CONVERSE, S. E. KAHLER, S. T. BARLOW, B. A. STONE, und R. S. BHOGAL: *The persona effect: Affective impact of animated pedagogical agents*. In *Proc. CHI*, S. 359–366. 1997.

[2] REEVES, B. und C. NASS: *The Media Equation. How People Treat Computers, Television and New Media Like Real People and Places*. Cambridge University Press, Cambridge, 1998.

[3] RAUCHBAUER, B., M. BOURHIS, M. OCHS, L. PRÉVOT, und T. CHAMINADE: *Brain activity during reciprocal social interaction investigated using conversational robots as control condition*. *Philosophical Transactions of the Royal Society B*, 374(1771), S. 20180033, 2019.

[4] ZLOTOWSKI, J., E. STRASSER, und C. BARTNECK: *Dimensions of anthropomorphism – from humanness to humanlikeness*. In *Proc. IEEE International Conference on Human-Robot Interaction*, S. 66–73. 2014.

[5] DENNETT, D. C.: *Intentional systems*. *The Journal of Philosophy*, 68(4), S. 87–106, 1971.

[6] MOORE, R. K.: *From talking and listening robots to intelligent communicative machines*. In J. MARKOWITZ (Hrsg.), *Robots That Talk and Listen*. De Gruyter, Boston, MA, 2015.

[7] VÖLKEL, S. T., C. SCHNEEGASS, und D. B. MALIN EIBAND: *What is "intelligent" in intelligent user interfaces? a meta-analysis of 25 years of IUI*. In *Proc. IUI*. 2020.

[8] GERO, K. I., Z. ASHKTORAB, C. DUGAN, Q. PAN, J. JOHNSON, W. GEYER, und M. R. ET AL.: *Mental models of ai agents in a cooperative game setting*. In *CHI Conference on Human Factors in Computing Systems (CHI)*, S. 1–12. 2020.

[9] BRENNAN, S. E.: *The grounding problem in conversations with and through computers*. In S. FUSSELL und R. KREUZ (Hrsg.), *Social and Cognitive Psychological Approaches to Interpersonal Communication*, S. 201–225. Lawrence Erlbaum, Hillsdale, NJ USA, 1998.

[10] ENFIELD, N.: *How we talk. The Inner Workings of Conversation*. BasicBooks, New York USA, 2017.

[11] KOUTSOMBOGERA, M. und C. VOGEL: *Speech pause patterns in collaborative dialogs*. In E. A. und J. L. (Hrsg.), *Innovations in Big Data Mining and Embedded Knowledge*, S. 99–115. Springer, Cham, 2019.

[12] KNAPP, M. und J. HALL: *Nonverbal Communication in Human Interaction*. Wadsworth, Boston, MA USA, 2010.

[13] TSENG, S.-H., Y.-H. HSU, Y. S. CHIANG, T.-Y. WU, und L.-C. FU: *Multi-human spatial social pattern understanding for a multi-modal robot through nonverbal social signals*. In *23rd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, S. 531–536. 2014.

[14] HONIG, S. und T. ORON-GILAD: *Understanding and resolving failures in human-robot interaction: Literature review and model development*. *Frontiers in Psychology*, 9, S. 861, 2018.

[15] MILLER, T.: *Explanation in artificial intelligence: Insights from the social sciences*. *Artificial intelligence*, 267, S. 1–38, 2019.

[16] SUH, J., C. C. BENNETT, B. WEISS, E. YUN, J. JEONG, und Y. CHAE: *Development of speech dialogue systems for social AI in cooperative game environments*. In *IEEE TENSYMP*, S. 1–4. 2021.

[17] SARAH T, V., R. SCHÖDEL, D. BUSCHEK, C. STACHL, V. WINTERHALTER, M. BÜHNER, und H. HUSSMANN: *Developing a personality model for speech-based conversational agents using the psycholexical approach*. In *CHI Conference on Human Factors in Computing Systems (CHI)*, S. 1–14. 2020.

[18] BARTNECK, C., E. CROFT, D. KULIC, und S. ZOGHBI: *Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots*. *International Journal of Social Robotics*, 1(1), S. 71–81, 2009.

[19] BIOCCA, F., C. HARMS, und J. GREGG: *The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity*. In *4th annual international workshop on presence*, S. 1–9. 2001.

[20] LEE, H. R. und S. SABANOVIC: *Culturally variable preferences for robot design and use in south korea, turkey, and the united states*. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, S. 17–24, 2014.