

SSML für humanoide Serviceroboter

Prof. Dr. **O. Bendel**, Hochschule für Wirtschaft FHNW, Basel

Postprint, 06.12.2017, publiziert über informationsethik.net und oliverbendel.net; gedruckt veröffentlicht am 05.12.2017: Bendel, Oliver. SSML für humanoide Roboter. In: Humanoide Roboter 2017. Tagungsband zur VDI-Konferenz „Humanoide Roboter“ vom 05. bis 06.12.2017 in Aschheim bei München. VDI Wissensforum, Aschheim 2017. Für Zitationen gilt die vorliegende Version.

Prof. Dr. oec. HSG Oliver Bendel
Fachhochschule Nordwestschweiz
Hochschule für Wirtschaft
Institut für Wirtschaftsinformatik
Bahnhofstrasse 6
CH-5210 Windisch

Kurzfassung

Humanoide Roboter sind oft sprechende Roboter. Zur Erzeugung und Manipulation der künstlichen Stimme gibt es verschiedene Ansätze. Umgesetzt wird die Sprachsynthese meist mit einem Text-to-Speech-System. VoiceXML und die Speech Synthesis Markup Language (SSML) sind wichtig bei der Gestaltung der Interaktion mit dem menschlichen Benutzer und der Imitation der menschlichen Stimme. Zunehmend werden auch automatische Lernverfahren eingesetzt. Bei SSML steht ein Satz an Befehlen zur Verfügung. Nun verbreiten sich die humanoiden Roboter, und mehr und mehr werden wir in Alltagssituationen mit ihnen konfrontiert. Man kann sie mit künstlichen Stimmen ausstatten und diese mit SSML gestalten. Die Frage ist, ob die Sprache in ihrem jetzigen Umfang für alle Bedürfnisse ausreicht oder ob es spezifische Befehle für spezifische Erscheinungsformen und Gelegenheiten braucht. Der Beitrag fokussiert auf humanoide Serviceroboter, schlägt eine Einteilung für diese nach Zwecken und Funktionen vor und untersucht jeweils den Gebrauch und die mögliche Erweiterung von SSML.

Abstract

Humanoid robots are often speaking robots. There are different approaches to the creation and manipulation of the artificial voice. Speech synthesis is usually implemented with a text-to-speech system. VoiceXML and Speech Synthesis Markup Language (SSML) are important in shaping interaction with the human user and imitating the human voice. Increasingly, automatic learning methods are also being used. SSML provides a set of commands. Now there

are more and more humanoid robots, and more and more we are confronted with them in everyday situations. They can be equipped with artificial voices and designed with SSML. The question is whether this language is sufficient for all needs in its current scope or whether specific commands for specific manifestations and occasions are needed. The article focuses on humanoid service robots, proposes a classification according to purposes and functions and examines the use and possible extension of SSML.

1. Einleitung

Humanoide Roboter sind oft sprechende Roboter. Dabei tönt es nicht nur aus den Maschinen, so wie die Stimme von Siri und Cortana aus den Smartphones und die Stimme von Alexa aus Echo und Echo Dot, sondern das Gesprochene geht einher mit Mimik und Gestik sowie mit Handlungen aller Art, und der Roboter reagiert auf (Sprech-)Handlungen des Gegenübers [3, 6, 8, 14]. Grundsätzlich sehen humanoide Roboter nicht neutral aus, sondern verkörpern häufig etwas, repräsentieren z.B. Männer oder Frauen, Erwachsene oder Kinder. Es ist insofern nicht gleichgültig, welche Stimme mit welchen Ausdrucksmöglichkeiten man ihnen verleiht. Generell spielt die Stimme eine große Rolle in sozialen Zusammenhängen [16].

Zur Erzeugung und Manipulation der künstlichen Stimme gibt es verschiedene Ansätze. Umgesetzt wird die Sprachsynthese meist mit einem Text-to-Speech-System [15]. VoiceXML und die Speech Synthesis Markup Language (SSML) sind wichtig bei der Gestaltung der Interaktion mit dem menschlichen Benutzer und der Imitation der menschlichen Stimme. Zunehmend werden automatische Lernverfahren eingesetzt. Bei SSML steht ein Satz an Befehlen zur Verfügung. Nun verbreiten sich die humanoiden Roboter, und mehr und mehr werden wir in Alltagssituationen mit ihnen konfrontiert. Man kann sie mit künstlichen Stimmen ausstatten und diese mit SSML gestalten [14]. Die Frage ist, ob die Sprache – wenn man sich für sie entscheidet – in ihrem jetzigen Umfang für alle Bedürfnisse ausreicht oder ob es spezifische Befehle für spezifische Erscheinungsformen und Gelegenheiten braucht.

Der vorliegende Beitrag fokussiert auf humanoide Serviceroboter, schlägt eine Einteilung für diese nach Zwecken und Funktionen vor und untersucht, ausgehend von geeigneten Typen (Informations- und Navigationsroboter, Unterhaltungs- und Spielzeugroboter, Pflege- und Therapieroboter), den Gebrauch und die Erweiterungsmöglichkeiten von SSML (die dem W3C unterbreitet werden könnten [19]). Die technische Realisierung wird nicht thematisiert.

2. Sprachsynthese und SSML

Künstliche Stimmen haben eine Tradition, die bis ins 19. Jahrhundert zurückreicht. Der Erfinder des Schachtürken etwa, Wolfgang von Kempelen, experimentierte in diesem Feld [10]. Während der berühmte Automat ein Fake war, was später Edgar Allan Poe in aller Klarheit darlegte [18], war der Versuch zur Stimmerzeugung von Erfolg gekrönt. Weitere Ingenieure und Tüftler wie Charles Wheatstone und Joseph Faber präsentierten Sprachautomaten, die ihre Bewunderer fanden.

Ab den 50er-Jahren versuchte man, den Computern das Sprechen beizubringen. Das erste computerbasierte Sprachsynthesesystem wurde in den späten 1950ern, das erste komplette TTS-System 1968 fertiggestellt [1, 11, 15]. Der Physiker John Larry Kelly, Jr entwickelte 1961 bei den Bell Labs eine Sprachsynthese mit einem IBM 704 und ließ ihn das Volkslied „Daisy Bell“ singen. Das heutige IBM Watson besitzt ebenfalls eine TTS-Engine, erreichbar über www.ibm.com/watson/developercloud.

Die Speech Synthesis Markup Language basiert auf XML [1, 19]. Wurzelement ist der Tag `<speech>`, der mit `</speech>` abgeschlossen wird (so wie `<html>` bzw. `</html>` bei der Hypertext Markup Language, kurz HTML, der Seitenbeschreibungssprache für das World Wide Web). Es existieren spezifische Elemente, etwa `<voice>`, die die Kategorie der Sprachsynthese angeben, zudem Attributnamen (im Folgenden Attribute genannt) und die Werte der Attribute (im Folgenden Werte genannt). Gültige Attribute von `<speech>` sind beispielsweise *version* (obligatorisch) und *xml:lang* (optional). Es treten in SSML ungefähr 40 Elemente oder Tags auf, die zur Veränderung und Anpassung einer synthetischen Stimme verwendet werden können [1, 19]. Anbieter wie IBM und Amazon unterstützen einige davon und haben zudem Erweiterungen für ihre Services (etwa die TTS von IBM Watson und Alexa für Echo) definiert.

Mit SSML-Anweisungen können Sätze, Wörter und Wortbestandteile manipuliert werden. Wie bei HTML entstehen teils komplexe Verschachtelungen. Die Struktur sieht bei SSML wie folgt aus (die Einrückung hat keine Funktion und dient lediglich der Übersichtlichkeit):

```
<tag1>
  <tag2 attributeName="attributeValue">
    Beispieltext
  </tag2>
</tag1>
```

Es gibt etablierte SSML-Befehle, mit denen man beispielsweise die Lautstärke, die Geschwindigkeit oder die Stimmhöhe manipuliert [1]. Bei der Engine von IBM Watson (bei der Expressive SSML und Voice Transformation SSML unterschieden werden) kann z.B. der Grad der Begeisterung dosiert werden. Es dürfen mehrere Attribute nebeneinanderstehen, jeweils konkretisiert durch Werte. Diese können, wie in HTML, sowohl Wörter als auch Zahlen sein.

Amazon hat im Jahre 2017 fünf SSML-Befehle eingeführt, um Alexa menschlicher klingen zu lassen [13]:

1. Whispers – Convey a softer dialog with `<amazon:effect name="whispered">`.
2. Expletive beeps – Bleep out words with `<say-as interpret-as="expletive">`.
3. Sub – Use the `<sub>` tag when you want Alexa to say something other than what's written.
4. Emphasis – Add `<emphasis>` to change the rate and volume at which Alexa speaks.
5. Prosody – Use this tage [!] to control the volume, pitch, and rate of speech.

Auf den Befehl, der das Flüstern ermöglicht, wird weiter unten eingegangen, ebenso auf künstlich produzierte Verunsicherung und Begeisterung.

Offensichtlich besteht Bedarf, virtuelle Stimmen zu modellieren und zu adaptieren. Bei humanoiden Robotern kann ein solcher ebenso angenommen werden, und sie sollen eben nicht nur an sich menschlich, sondern menschlich im Einklang mit ihrem menschlichen Äußeren und ihren Zielen und Funktionen klingen, zudem im Zusammenspiel mit Mimik und Gestik. Natürlich sind virtuelle Stimmen auch bei nichthumanoiden Robotern und sogar bei simplen Automaten und Geräten sinnvoll, und dort sind sie womöglich sogar dasjenige Merkmal, das die Maschine in der Interaktion mit dem Menschen wesentlich prägt. Allerdings ergeben sich bei Autos und Navis, um nur zwei Beispiele zu nennen, zum Teil andere Anforderungen und Fragestellungen.

3. SSML für humanoide Serviceroboter

Im Folgenden eine knappe Auswahl von Tags, die im vorliegenden Kontext des Einsatzes von Servicerobotern relevant sind, nebst einer kurzen Einschätzung [1]. Gefolgt wird Version 1.1 der W3C-Empfehlung vom 7. September 2010 [19]; dargestellt werden zudem Entwicklungen von IBM (die auf Version 1.0 basieren) und Amazon.

<voice>

Dieser Tag wird gebraucht, um grundsätzliche Merkmale einer Stimme zu beeinflussen. Mögliche Attribute sind *age* (es sind Zahlenwerte zulässig), *gender* (zulässige Werte sind "male", "female" und "neutral") und *name* (angegeben werden die Namen der installierten Stimmen). Damit können Identität und Persönlichkeit des Sprechers modelliert werden. Da Alter und Geschlecht durchaus Kategorien sind, die bei der Vermittlung von Dienstleistungen und Produkten relevant sind, ist eine Verwendung für Serviceroboter angezeigt. So sollte in vielen (nicht in allen) Zusammenhängen ein Serviceroboter nicht zu jung klingen (und nicht zu jung aussehen), auf jeden Fall nicht wie ein Kind. Doch auch eine zu alte Stimme könnte den Aufbau von Vertrauen – um den es hier im Kern geht – belasten. IBM und Amazon unterstützen dieses Element nicht bzw. nicht in vollem Umfang – die Stimmen, die zur Verfügung stehen, gehören zu fiktionalen Männern und Frauen eines bestimmten Alters (allerdings können sie, wie der nächste Absatz zeigt, verjüngt werden).

<voice-transformation>

<voice-transformation> wird von IBM Watson umgesetzt, mit dem Attribut *type*. Mögliche Werte sind "Young" (für eine jüngere Stimme) und "Soft" (für eine weichere Stimme). Die Stärke von "Young" und "Soft" kann durch das zusätzliche Attribut *strength* reguliert werden (Angaben in Prozent). Mit einem weiteren Wert namens "Custom" werden die Attribute *glottal_tension* (Angaben in Prozent), *breathiness* (Angaben in Prozent) und *timbre* (mögliche Werte sind *Sunrise* und *Breeze*) verbunden. Eine zu junge Stimme kann, wie gesagt, bei Servicerobotern problematisch sein, zumal im Zusammenhang mit einem jungen Äußeren. Eine weiche Stimme oder ein hingehauchtes Wort können hingegen beim Verkauf und bei der Informierung von Vorteil sein. Auch Menschen nehmen sich in solchen Vorgängen zurück und vermeiden tendenziell eine zu harte Betonung und einen zu strengen Ausdruck.

<express as>

<express as> wird von IBM Watson unterstützt. Als Attribut wird *type* verwendet. Mögliche Werte sind "Apology", "Uncertainty" und "GoodNews". Der Ausdruck des Bedauerns scheint für den humanoiden Serviceroboter wichtig zu sein. Wenn seine Dienstleistungen nicht perfekt sind oder er Informationen nicht liefern kann, kann er dies entsprechend begleiten und ausgleichen. Wenn eine Frage, eine Aussage bzw. ein Sachverhalt nicht genau beurteilt werden können, mag der Ausdruck der Unsicherheit passend sein. Wird für ein Produkt geworben, hat der Mensch ein Ziel erreicht oder wird der Roboter gelobt, ist der "GoodNews"-Wert von Vorteil. Der humanoide Serviceroboter kann jeweils das Gesprochene mit Bewegungen, Gesten und Mimik begleiten.

<prosody>

<prosody> wird von IBM Watson und Amazon unterstützt. Man kann *volume* (gültige Werte zwischen 1 und 100), *pitch* (verschiedene mögliche Werte, u.a. Angaben in Prozent) und *rate* (verschiedene mögliche Werte, u.a. Anzahl der Worte pro Minute und Angaben in Prozent) als Attribute einsetzen. Lautstärke und Tonhöhe haben für einen Serviceroboter durchaus Bedeutung. Dieser kann sich, wenn dies gewünscht und passend ist, in seiner Lautstärke auf sein Gegenüber einstellen. Wenn er äußerlich als Mann oder Frau gestaltet ist, kann er sich je nach Geschlecht unterschiedlich artikulieren. Frauen schätzen bei Männern eher eine tiefe Stimme, Männer bei Frauen eher eine hohe. Für den Verkaufserfolg ist dies ebenso wichtig wie für das Wohlbefinden.

<amazon:effect>

<amazon:effect> mit dem Attribut *name* wurde von Amazon mit dem Wert "whispered" belegt. Damit ist nicht nur leises Sprechen möglich, sondern die Stimme wird gleichsam zurückgenommen bzw. ausgeschaltet. Alexa ist sozusagen ein virtueller Serviceroboter. Offensichtlich besteht Interesse daran, sie in dieser Weise sprechen zu lassen, obwohl sie meist in einem privaten Umfeld zugange ist – laut Amazon dienen die Erweiterungen einfach dazu, die Stimme menschenähnlicher zu machen. In einem öffentlichen Umfeld kann das Flüstern benutzt werden, um jemanden diskret zu informieren, etwa im Falle interaktiver Werbeflächen mit personalisierten Funktionen [2]. Auch humanoide Roboter sind mögliche Anwendungsfälle, und sie können das Flüstern mit entsprechenden Bewegungen und Gesten (Neigen des Kopfes etc.) begleiten, sodass der Benutzer vorbereitet und informiert ist.

<emphasis>

<emphasis>, ein in SSML vorgesehener Tag, wurde von Amazon für Alexa für das Attribut *level* mit den Werten "strong", "moderate" und "reduced" versehen [13]. Hier sind spezielle Eignungen für die unterschiedlichen Situationen möglich, in denen humanoide Serviceroboter und Benutzer interagieren und kommunizieren, und gerade zusammen mit anderen Elementen dürfte es interessante Anwendungsmöglichkeiten geben.

<say-as>

Der Tag <say-as> in Kombination mit dem Attribut *interpret-as* ist wichtig, um Zahlen und Wörter korrekt bzw. nach den bestehenden Konventionen auszusprechen. Der Wert "digits" etwa erzwingt, dass Zahlen, der Wert "letters", dass Buchstaben einzeln vorgelesen werden. Hu-

manoide Serviceroboter, gerade solche, die präzise und verständliche Informationen übermitteln sollen, sind auf Möglichkeiten dieser Art angewiesen. Der genannte Wert "expletive" kann ebenfalls nützlich sein.

Deutlich wurde, dass mit dem Standardrepertoire von SSML bereits einige der gewünschten Anpassungen und Wirkungen erzeugt werden können. Sicherlich fehlen aber ein paar konkrete Möglichkeiten in den identifizierten Bereichen. Auf diese wird im nächsten Abschnitt eingegangen. Eine technische Spezifikation findet, wie bereits angedeutet, nicht statt.

4. SSML für humanoide Serviceroboter

Serviceroboter sind für Dienstleistung, Unterhaltung und Zuwendung zuständig, sie holen Geschirr und Besteck, Nahrungsmittel und Medikamente, überwachen die Umgebung ihrer Besitzer oder den Zustand von Patienten und halten ihr Umfeld im gewünschten Zustand [4, 5]. Wenn sie mit Sensoren ausgestattet sind, wenn sie über Intelligenz und Erinnerungsvermögen verfügen, werden sie nach und nach zu allwissenden Begleitern. Sie wissen, was ihr Besitzer tut, was er sagt, wie er sich fühlt, was er trägt, was in seiner Umwelt geschieht, und sie wissen, was die Passanten in den Städten und Dörfern berührt, und melden es womöglich an ihre Betreiber oder an Geräte und Computer aller Art.

So wie die Industrieroboter immer mehr ihre Gefängnisse und Behausungen verlassen, so wie sie immer mobiler und universeller geraten, und so wie sie immer mehr an den Menschen heranrücken – man denke an Kooperations- und Kollaborationsroboter, kurz Co-Robots –, so befreien sich die Serviceroboter immer mehr von inneren und äußeren Zwängen und treten in eine Koexistenz mit uns ein [4, 5]. Serviceroboter können in folgende Kategorien eingeteilt werden [5]:

- Haushalts- und Gartenroboter
- Sicherheits- und Überwachungsroboter
- Transport- und Lieferroboter
- Informations- und Navigationsroboter
- Unterhaltungs- und Spielzeugroboter
- Pflege- und Therapieroboter

Auch Militär- und Kampfroboter sowie Weltraumroboter werden zuweilen zu den Servicerobotern gezählt, zudem Sexroboter, von denen es freilich nur wenige Umsetzungen gibt [1]. Nicht

alle der aufgelisteten Roboter werden humanoid gestaltet. Bei den drei zuletzt genannten Typen sind aber jeweils mehrere Beispiele dafür vorhanden. Sie werden im vorliegenden Beitrag näher untersucht. Ausgehend von Gestaltung, Zweck und Funktionen sowie einer Beschreibung typischer Situationen wird gefragt, welche älteren und neueren Tags, Attribute und Werte in SSML wichtig für die Information von und Kommunikation mit Benutzern sind. Zudem wird untersucht, welche Tags, Attribute oder Werte benötigt werden, um die Serviceroboter als humanoide Roboter ihrer Aufgabe gerecht werden zu lassen. Die neuen Anweisungen sollen also einerseits die Menschenähnlichkeit der Roboter, andererseits ihre Alltagstauglichkeit unterstützen.

Informations- und Navigationsroboter

Informations- und Navigationsroboter fahren oder gehen über Parks und Gelände, durch Museen, Messen und Verkaufsräume und informieren Besucher und Kunden über Veranstaltungen und Möglichkeiten der Besichtigung und führen sie an die gewünschte Stelle [5]. Nicht zuletzt sind sie in Hotels zu sehen, etwa an der Rezeption [12]. Sie verfügen häufig über Displays und spezielle Touchscreens und über natürlichsprachliche Fähigkeiten. Entsprechend unterbreiten sie textuelle respektive visuelle Informationen oder unterhalten sich mit den Benutzern. Zudem verwenden sie Karten und Technologien wie GPS.

Ein Beispiel ist Obelix, der in Städten als Fremdenführer fungieren soll, ein Prototyp der Universität Freiburg, der ETH Zürich und anderer Forschungs- und Entwicklungseinrichtungen, so groß wie ein Jugendlicher [7]. Immer mehr werden persönliche bzw. emotionale Roboter wie Pepper [9, 17] (eigentlich ein Companion- bzw. Unterhaltungs- und Spielzeugroboter) und Care-O-bot (eigentlich ein Pflegeroboter bzw. ein kombinierter Informations- und Navigationsroboter und Transport- und Lieferroboter, der für die Pflege gedacht ist) sozusagen umfunktionierte und in Shopping Malls und auf Messen eingesetzt.

Bei Informations- und Navigationsrobotern ist die differenzierte und korrekte Aussprache von Zahlen, Formeln etc. wichtig. Dafür stehen in SSML über `<say-as>` in Kombination mit dem Attribut *interpret-as* verschiedene Werte zur Verfügung. Ebenso von Bedeutung ist das erwähnte `<express as>` mit dem Attribut *type* und den Werten "Apology", "Uncertainty" und "GoodNews". Damit kann etwa in Shopping Malls, in Verkaufsräumen sowie in Museen und auf Messen entsprechend kommuniziert werden, wenn Erwartungen nicht, nicht ganz oder voll erfüllt werden.

Innerhalb von `<express as>` mit dem Attribut *type* wären weitere Werte wie "Certainty" (als Gegenteil von "Uncertainty", um Bestimmtheit auszudrücken, etwa wenn der Kunde versucht, den Preis zu drücken; ein ähnlicher Effekt lässt sich womöglich mit *glottal_tension* erreichen) und "BadNews" (als Gegenteil von "GoodNews", um noch mehr als bei "Apology" das Ungute der Situation anzudeuten, etwa wenn sich ein Hurrikan nähert und der Roboter die Aufgabe hat, dies mitzuteilen und zu illustrieren) interessant. Nicht zuletzt könnte `<voice-transformation>` mit dem Attribut *type* um den Wert "Old" ergänzt werden, um eine künstliche Stimme künstlich altern zu lassen und damit (ähnlich wie mit einem hohen Wert bei *age* innerhalb von `<voice>`) eine gewisse Seriosität herzustellen.

Unterhaltungs- und Spielzeugroboter

Unterhaltungs- und Spielzeugroboter dienen der Unterhaltung und Zerstreuung von Benutzern, nicht nur von Kindern und Jugendlichen, sondern auch von Erwachsenen [5]. Auch zum Lernen kann man manche von ihnen verwenden. Sie tanzen, singen, spielen Musik, erlauben ihre Konstruktion und Dekonstruktion. Manche sind menschen-, andere tierähnlich. Einige sind handgroß, andere haben die Dimensionen von Katzen und Hunden, denen sie nachgebildet sein können, wieder andere von Kindern und Jugendlichen.

WowWee und Splash Toys bieten über diverse Onlinehändler elektronische Spielzeughunde an [5]. Die LEGO-Roboter (LEGO MINDSTORMS) nehmen je nach Kreativität der Benutzer ganz unterschiedliche Gestalt an. Nao von Aldebaran bzw. SoftBank kann recht universell gebraucht werden, ebenso der bereits erwähnte Pepper aus dem gleichen Haus.

Bei Unterhaltungs- und Spielzeugrobotern ist die Verwendung `<voice>` mit dem Attribut *age* und Zahlenwerten zwischen 4 und 14 durchaus sinnvoll. Kinder bevorzugen nicht durchweg das Spiel mit künstlichen Gleichaltrigen, wie Babypuppen sowie Barbie und Ken zeigen, aber gerade bei humanoiden Robotern könnte eine gewisse Präferenz existieren. Auch `<amazon:effect>` mit dem Attribut *name* und dem Wert "whispered" ist interessant, beispielsweise wenn Geheimnisse erzählt werden sollen.

Bei Robotern für Kleinkinder wären neue Anweisungen interessant, die Babytalk als Resultat haben, etwa der Wert "Baby" innerhalb von `<voice-transformation>` mit dem Attribut *type*. Weiter könnte man Robotern zumindest Sprechgesang ermöglichen. Es ist nicht trivial, einem System Formen des Singens beizubringen, und SSML dürfte nur ein Stück weit dafür geeignet

sein. Dennoch gibt es hier Möglichkeiten, etwa im Zusammenhang mit <prosody>, neuen Werten für Attribute wie *volume* und *pitch* und neuen Attributen. Mimik und Gestik können noch unbeholfene Versuche des Serviceroboters ein Stück weit ausgleichen.

Der Ausdruck von Emotionen kann durch SSML ebenfalls besser unterstützt werden. Zu denken wäre an Elemente und Attribute, die das Lachen regeln, beispielsweise innerhalb von <voice-transformation>. Manipuliert würden dabei Wörter wie "hahaha" oder "hehehe". Es geht darum, dass sie nicht wie vorgelesen, sondern wie ausgestoßen klingen. Wenn der humanoide Roboter dabei seinen Mund verzieht oder sich auf die Schenkel klopft, kann das die Wirkung steigern.

Pflege- und Therapieroboter

Pflegeroboter unterstützen oder ersetzen menschliche Pflegekräfte. Sie bringen und reichen Kranken und Alten benötigte Medikamente und Nahrungsmittel, helfen ihnen beim Hinlegen und Aufrichten oder alarmieren Angehörige oder einen Notdienst [4]. Manche verfügen über natürlichsprachliche Fähigkeiten, sind lernende und intelligente Systeme.

Beispiele für Entwicklungen und Prototypen von Pflegerobotern sind JACO, Care-O-bot, Cody und Robear (Vorgängerversionen RIBA und RIBA-II). JACO, ein Arm samt Hand mit drei Fingern, kann alles in Griffnähe besorgen, Care-O-bot, ein mobiler Assistent, sogar alles aus dem Nebenraum. Cody wäscht ans Bett gefesselte Patienten. Robear, der an einen Teddy erinnert, hebt sie hoch und lagert sie um, zusammen mit einer Pflegekraft.

Therapieroboter unterstützen therapeutische Maßnahmen oder wenden selbst, häufig als autonome Maschinen, solche an [4]. Sie sind mit ihrem Aussehen und in ihrer Körperlichkeit wie traditionelle Therapiegeräte präsent, machen aber darüber hinaus selbst Übungen mit Gelähmten, unterhalten Betagte und fordern Demente und Autisten mit Fragen und Spielen heraus. Manche verfügen über mimische, gestische und sprachliche Fähigkeiten und sind in einem bestimmten Umfang denk- und lernfähig.

Wohlbekannt selbst bei nicht betroffenen Personen und Gruppen ist die Kunstrobbe Paro (www.parorobots.com), die seit Jahren im Einsatz ist. Ebenfalls bekannt ist Keepon (www.my-keepon.com), ein kleiner, gelber Roboter, der die soziale Interaktion von autistischen Kindern beobachten und verbessern soll und inzwischen auf dem Massenmarkt erhältlich ist.

Von den oben erwähnten Optionen scheint u.a. <voice-transformation> mit dem Attribut *type* und dem Wert "Soft" relevant, zudem <amazon:effect> mit dem Attribut *name* und dem Wert "whispered", um Zuwendung, Anteilnahme oder Rücksichtnahme gegenüber Patientinnen und Patienten auszudrücken.

Für Pflegeroboter wäre wiederum der Wert "Baby" interessant („Wie geht es uns denn heute?“), innerhalb von <voice-transformation> mit dem Attribut *type*. Man darf diese Sprechweise aber durchaus kritisieren und darauf verzichten, sie zu implementieren. Zudem wird innerhalb von <voice> das Attribut *style* mit dem Wert "robot" vorgeschlagen. Wenn man die Stimme gerade nicht natürlich gestalten will, etwa um eine zu enge Bindung zwischen Mensch und Maschine zu vermeiden, kann man sie roboterhaft klingen lassen, wie dies bei älteren Systemen selbstverständlich und bei neueren Maschinen wie Pepper absichtsvoll vorkommt. Die Idee dabei ist, die Bindung zwischen Pflegeroboter und -bedürftigem nicht zu eng werden zu lassen, um den Menschen sowohl Vortäuschungen als auch Enttäuschungen zu ersparen.

Für manche Therapieroboter kann der zuletzt genannte Befehl ebenfalls interessant sein. Allerdings kann bei Erkrankungen eine Illusionierung angezeigt sein, zumindest eine glaubwürdig klingende Zuwendung. Bei <express-as> wird innerhalb von *type* etwa "tender" in Betracht gezogen. Mit Zärtlichkeit in der Stimme kann ein Gefühl gegenüber dem menschlichen Partner ausgedrückt und dieser kurzfristig und dauerhaft gebunden werden. Ein Stück weit mag hier auch *breathiness* helfen.

5. Zusammenfassung und Ausblick

Der Beitrag widmete sich der Sprachsynthese für humanoide Serviceroboter, unter besonderer Berücksichtigung von SSML und mit Blick auf den konkreten Einsatz. Das Ziel war die Auswahl bestehender und der Vorschlag neuer Tags, Attribute und Werte, die bei Bedarf technisch umgesetzt werden können.

Ein anspruchsvolles Gebiet, das hier nicht weiter behandelt werden konnte, ist die technische Integration von Stimme, Mimik und Gestik. Während man bei Avataren und Softwarerobotern diesbezüglich schon weit gekommen ist, ist die Lage bei Hardwarerobotern anders. Nur wenige verfügen überhaupt über ausgefeilte mimische und gestische Fähigkeiten.

- [1] Bendel, O.: SSML for Sex Robots. In: Cheek, A. D.; Devlin, K.; Levy, D. (Hrsg.): Love and Sex with Robots. Third International Conference, LSR 2017, London, UK, 19. – 20. Dezember 2017, Revised Selected Papers. Cham: Springer International Publishing 2018 (erscheint 2018)
- [2] Bendel, O.: Neue Spione in den Straßen, auf den Plätzen und in den Läden: Interaktive Werbeflächen aus ethischer Sicht. Telepolis, 15. August 2017.
<https://www.heise.de/tp/features/Neue-Spione-in-den-Strassen-auf-den-Plaetzen-und-in-den-Laeden-3797118.html>
- [3] Bendel, O.: The Synthetization of Human Voices. AI & SOCIETY, 26. Juli 2017 („Online-First“-Artikel auf SpringerLink). <http://rdcu.be/uvxm>
- [4] Bendel, O.: Serviceroboter. Beitrag für das Gabler Wirtschaftslexikon. Springer Gabler, Wiesbaden 2017. <http://wirtschaftslexikon.gabler.de/Definition/serviceroboter.html>
- [5] Bendel, O.: Service Robots in Public Spaces. Telepolis, 25. Juni 2017.
<https://www.heise.de/tp/features/Service-Robots-in-Public-Spaces-3754173.html>
- [6] Bendel, O.: Der kleine Lauschangriff: Auditive Systeme aus Sicht der Ethik. Telepolis, 5. Juli 2015. <http://www.heise.de/tp/artikel/45/45319/1.html>
- [7] Burgard, W.; Stachniss, C.: Gestatten, Obelix. forschung 1/2013 S. 4–8
- [8] Caglayan, A. K.; Harrison, C. G.: Intelligente Software-Agenten: Grundlagen, Technik und praktische Anwendung im Unternehmen. München: Hanser 1998
- [9] Coeckelbergh, M.: Personal Robots, Appearance, and Human Good: A Methodological Reflection on Roboethics. International Journal of Social Robotics, August 2009, Volume 1, Issue 3 S. 217–221
- [10] Kempelen, W. v.: Mechanismus der menschlichen Sprache nebst der Beschreibung seiner sprechenden Maschine. Wien: J. B. Degen 1791
- [11] Klatt, D.: Review of text-to-speech conversion for English. J. Acous. Soc. Amer. 82, 1987 S. 737–793
- [12] Lill, F.: „Sprechen Sie deutlich!“ ZEIT ONLINE, 10. September 2015.
<http://www.zeit.de/2015/35/roboter-henn-na-hotel-japan-nagasaki>
- [13] Myers, L.: New SSML Features Give Alexa a Wider Range of Natural Expression. Blog der Amazon Developer Website, 27. April 2017. <https://developer.amazon.com/de/blogs/alexa/post/5c631c3c-0d35-483f-b226-83dd98def117/new-ssml-features-give-alexa-a-wider-range-of-natural-expression>
- [14] Roehling, S.; MacDonald, B.; Watson, C.: Towards Expressive Speech Synthesis in English on a Robotic Platform. In: Warren, P.; Watson, C. I.: Proceedings of the 11th

Australian International Conference on Speech Science & Technology, 6. – 8. Dezember 2006 S. 130–135

- [15] Taylor, P.: Text-to-Speech Synthesis. Cambridge: Cambridge University Press 2009
- [16] Wachsmuth, I.: Menschen, Tiere und Max: Natürliche Kommunikation und künstliche Intelligenz. Heidelberg: Springer Spektrum 2012
- [17] Wendel, J.: Pepper The Robot soll nicht für Sex benutzt werden. Wired, 24. September 2015. <https://www.wired.de/collection/latest/eine-passage-im-nutzervertrag-von-pepper-robot-verbietet-sex>
- [18] Wimsatt, W. K.: Poe and the Chess Automaton. In: On Poe: The Best of „American Literature“. Durham, NC: Duke University Press 1993
- [19] W3C (Hrsg.): Speech Synthesis Markup Language (SSML) Version 1.1. W3C Recommendation, 7. September 2010. <https://www.w3.org/TR/speech-synthesis11/>