

# MEHR MENSCH ALS MASCHINE?



## ZUR NATÜRLICHKEIT VON SPRACHDIALOGSYSTEMEN

Wenn es um das Design von Sprachdialogsystemen geht, wird häufig gefordert, diese möglichst „natürlich“ und „menschlich“ zu gestalten. Denn - so wird angenommen - nur ein besonders menschliches System sei auch ein gutes System. Aber ist das wirklich der Fall? Oder kann Menschlichkeit auch zum Nachteil werden? Und was bedeutet „Menschlichkeit“ überhaupt?

### Menschlichkeit - was ist das?

Hierfür zunächst ein Beispiel einer Mensch-Mensch-Kommunikation:

Agent	<i>Hallo, willkommen bei Payback. Was kann ich für Sie tun?</i>
Anrufer	<i>(verärgert) Ja, hallo. Mit meinem PayBack-Konto stimmt was nicht. Die neuen Punkte sind da noch nicht drauf.</i>
Agent	<i>(ernsthaft) Hmm, das werde ich sofort über-prüfen. Haben Sie Ihre Kundennummer zur Hand?</i>
Anrufer	<i>Nee, hab ich nicht.</i>
Agent	<i>Kein Problem, dann machen wir das anders. Wie ist Ihr Name?</i>
Anrufer	<i>Peter Meier.</i>
Agent	<i>Und Ihr Geburtsdatum?</i>
Anrufer	<i>15.12.1975</i>
Agent	<i>Ok, ah ja: Da ist tatsächlich ein Buchungsfehler passiert. Das Ganze ist allerdings inzwischen behoben. Sie sollten die Punkte spätestens mor-gen gutgeschrieben bekommen.</i>
Anrufer	<i>Aha.</i>
Agent	<i>Tut mir leid für die Unannehmlichkeiten, aber das kann manchmal pas-sieren, weil wir auch von Unterhändlern abhängig sind und...</i>
Anrufer	<i>(noch verstimmt, aber wesentlich entspannter) Schon gut, danke, ist ja nicht Ihr Fehler. Ich weiß dann ja bescheid.</i>
Agent	<i>(wieder gelöster) Alles klar. Und wenn Sie noch Fragen haben, lassen Sie sich am besten direkt zu mir durchstellen. Mein Name ist Sven Plöger.</i>
Anrufer	<i>Gut, mach ich. Vielen Dank erstmal. Und auf Wiederhören.</i>
Agent	<i>Auf Wiederhören.</i>

Augenscheinlichste Eigenschaft des Menschen ist es, in Gesprächen flexibel zu sein. Menschen gehen aufeinander ein, indem Sie ein mentales Bild entwerfen - von der Situation und von ihrem jeweiligen Gesprächspartner. Und dieses mentale Bild ändert sich fortlaufend während des Gesprächs. Es wird maßgeblich beeinflusst durch folgende Faktoren:

**Wer ruft an?** *Ist es ein Neukunde, ein Premiumkunde, ein Beschwerdeanruf oder ein Mitarbeiter?*

**In welchem emotionalen Zustand ist der Anrufer?** *Neutral, fröhlich, verärgert?*

**Welches Vorwissen hat der Anrufer?** *Weiß er genau, welche Informationen er benötigt?*

**Welche Ziele hat der Anrufer** - *und mithilfe welcher Strategie kann man sie erreichen?*

**Wie verlief der Interaktionsdiskurs bisher?** *Hat der Anrufer schon einmal angerufen? Welches Wissen sammelt sich über das Gespräch hinweg an? Welches Bild entwirft der Anrufer vom Agenten und dem Unternehmen?*

Im Beispiel oben erkennt der Agent, dass der Anrufer verärgert ist. Sofort passt er sich der Situation an: Er wechselt vom aufgeschlossen-freundlichen Ton hin zu einem konzentriert-sachlichen. Er signalisiert, dass er den Anrufer ernst nimmt und alles tun wird, ihm zu entsprechen. Daraufhin ändert sich auch der emotionale Zustand des Anrufers.

Auch in der Zielführung zeigt sich Flexibilität: Das Standard-Login-Verfahren (per Kundennummer) funktioniert nicht und sofort weicht der Agent auf eine alternative Strategie aus.

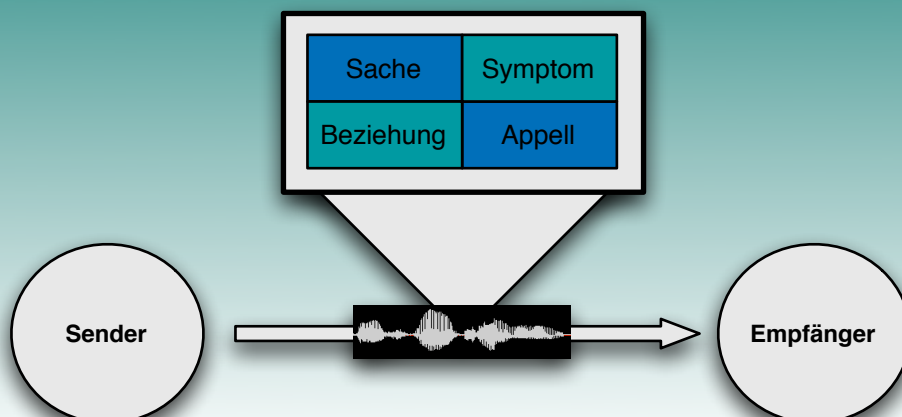
Die genannten Faktoren haben also einen Einfluss auf die einzelnen Äußerungen der Gesprächspartner. Jede der Äußerungen lässt sich sehr anschaulich durch das Vier-Seiten-Modell von Friedemann Schulz von Thun beschreiben, das auf das Organon-Modell von Karl Bühler zurückgeht:

Beispiel:

Anrufer	(verärgert) Ja, hallo. Mit meinem PayBack-Konto stimmt was nicht. Die neuen Punkte sind da noch nicht drauf.
Sache	Die neuen Punkte wurden noch nicht gutgeschrieben.
Beziehung	Ich bezahle Sie, damit Sie funktionieren.
Symptom	Ich bin verärgert.
Appell	Lösen Sie mein Problem!

Agent	(ernsthaft) Hmm, das werde ich sofort überprüfen. Haben Sie Ihre Kundennummer zur Hand?
Sache	Ich werde das Problem überprüfen.
Beziehung	Ich werde das erfüllen, was Sie von mir erwarten (zu funktionieren).
Symptom	Ich nehme das Problem ernst und bin daran interessiert, so schnell wie möglich zu einer Lösung zu kommen.
Appell	Geben Sie mir etwas Zeit und helfen Sie mir mit den notwendigen Informationen (Kundennummer).

Diese vier Seiten einer Nachricht zeigen sich auf allen Ebenen der Sprache: Der phonetischen, der syntaktischen und der lexikalischen. Dabei sind manche Ebenen der Sprache für bestimmte Seiten besser geeignet. Zum Beispiel wird die Symptomseite in erster Linie über die Stimmqualität (Phonetik) transportiert, während die sachliche Seite hauptsächlich über die Worte und die Syntax transportiert wird.



Nicht immer sind diese Seiten so homogen aufeinander abgestimmt wie hier. Zum Beispiel könnte der Agent einerseits mitteilen, dass er sich des Problems annimmt (Sachliche Seite), gleichzeitig aber über die Stimmqualität signalisieren, dass er es für lächerlich hält, sich deswegen so aufzuregen (Symptomseite, Beziehungsseite). Nicht selten geben unglücklich formulierte Äußerungen Anlass zu Missverständnissen, weil vor allem Beziehungs- und Symptomseite falsch interpretiert werden. In der alltäglichen Kommunikation sind diese Seiten sehr viel differenzierter als im Beispieldialog - denn die Gesprächsteilnehmer dort kennen sich nicht persönlich. Außerdem bewegt sich der Dialog in einem klar abgegrenzten thematischen Gebiet.

Die beschriebenen vier Seiten sind in jeder Nachricht vorhanden, aber sie sind nicht immer gleich stark ausgeprägt. Wissenschaftliche Texte und Dichtung sind Extrembeispiele: In wissenschaftlichen Texten findet sich fast ausschließlich die sachliche Seite, in der Dichtung ist die Symptomseite sehr stark ausgeprägt.

## Menschlichkeit in Sprachdialogsystemen

Selbst, wenn ein Sprachdialogsystem nicht die oben beschriebene Flexibilität aufweist, ist alleine schon deswegen ein menschliches Element vorhanden, weil die Interaktion per gesprochener Sprache stattfindet. Wenn also die Rede davon ist, ein Sprachdialogsystem menschlich zu gestalten, ist gemeint, ein System bewusst mit noch mehr menschlichen Attributen zu versehen. Analog dazu ist mit „ein Dialogsystem bewusst nicht menschlich zu gestalten“ gemeint „ein Dialogsystem so wenig menschlich wie möglich zu gestalten“.

Was steckt hinter der Forderung, Sprachdialogsysteme möglichst menschlich zu gestalten?

Wahrscheinlich hat dies seine Ursache darin, dass kommerzielle Dialogsysteme noch immer auf geringe Akzeptanz stoßen und Anrufer sich schlecht betreut fühlen. Es wird angenommen, man könne dem mit einem „menschlicheren“ System entgegenwirken. Aber eignet sich Menschlichkeit wirklich dazu, der geringen Akzeptanz entgegenzuwirken?

Eine extreme Ansicht dazu vertritt Clifford Nass (2006, „Wired for Speech“):

*Humans use the same parts of the brain to listen to computers as they listen and speak to other humans. This commonality in the ‚hardware‘ and mechanisms in the brain underlies the users‘ social responses to speech technologies. Recent research has provided clear empirical evidence that users exhibit the same social and human-directed responses to speech interfaces.*

Dialog-Design sollte demnach also so menschlich wie möglich sein - denn der Mensch neigt dazu, Sprachgewohnheiten von der Mensch-Mensch auf die Mensch-Maschine-Kommunikation zu übertragen. Und wenn der Nutzer sein gewohntes Gesprächsverhalten 1:1 auf Sprachdialogsysteme übertragen kann, wäre der Lernaufwand gleich null.

Allerdings ist es Fakt, dass die Sprachtechnologie bei weitem nicht an die Fähigkeiten des Menschen heranreicht. Deswegen kann es im Laufe eines Dialogs immer wieder zu unnatürlichen Gesprächssituationen kommen. Man denke an Situationen, in denen Nutzeräußerungen wiederholt nicht verstanden werden, in denen Listen von phonetisch ähnlichen Einträgen ausgegeben werden oder in denen vom Anrufer geäußerte Kundennummern zurückgelesen werden.

Hier kommt die andere extreme Ansicht ins Spiel, beispielsweise vertreten durch Bruce Ballentine (2007, „It's Better to Be a Good Machine Than a Bad Person“).

Laut Ballentine sollten sich Sprachdialogsysteme so weit wie möglich von Menschen unterscheiden. Denn so können die Nutzer am besten einschätzen, was sie dem System an Eingaben (d.h. Äußerungen) zumuten können und was nicht. Ballantine betont, dass Menschlichkeit und Nutzerfreundlichkeit nicht zwangsläufig zusammengehören: Ein System mit geringer Menschlichkeit kann durchaus sehr nutzerfreundlich sein und umgekehrt. Man denke hierbei zum Vergleich an GUIs: Durch die Interaktion mit Bildschirm, Maus und Tastatur (statt gesprochener Sprache) entbehren diese Schnittstellen jeglicher Menschlichkeit. Dennoch gibt es hinsichtlich Nutzerfreundlichkeit und intuitiver Bedienbarkeit sehr gute aber auch sehr schlechte GUIs.

## Menschlichkeit, die sich lohnt

Wie so häufig haben beide Parteien recht. Menschliche Elemente können die Bedienung erleichtern und das Hear&Feel des Dialogs attraktiver gestalten. An der falschen Stelle eingesetzt, wirkt Menschlichkeit aber ungläubwürdig und kann dazu führen, dass der Dialog misslingt, weil der Nutzer nicht abschätzen kann, wie das System zu steuern ist.

Die Frage ist also, wo genau das Optimum zwischen den beiden Extrempositionen liegt.

Welche menschlichen Elemente lassen sich wie einsetzen, um den Dialog zu bereichern?

## Themen- und Strategiewechsel antizipieren

Ein System mit gutem Dialog-Design ist dem Gedankengang des Nutzers immer einen Schritt voraus. Dies kann im Rahmen dynamischer Adaption oder aber auch im Rahmen eines starren Systems geschehen. In jedem Fall sind gründliche Nutzeranalysen erforderlich, die in ein gut durchdachtes Design münden. Hier ein Beispiel zum Themenwechsel:

System	Sie haben eine neue Nachricht - jetzt anhören?
Anrufer	Ja, bitte.
System	Nachricht von heute, 17 Uhr 23: <Nachricht wird abgespielt> Soweit die Nachricht. Möchten Sie direkt antworten?
Anrufer	Ja.

Anstatt den Anrufer in ein Hauptmenü zu geleiten, wo er aus „Nachricht anhören“, „Nachricht versenden“ etc. wählen könnte, werden anhand der aktuellen Situation (neue Nachricht vorhanden) veränderte Annahmen über die Ziele des Anrufers getroffen und der Dialog entsprechend darauf angepasst. Die Folge ist ein wesentlich kürzerer Dialog und eine angenehmere weil intuitivere Navigation.

## Diskurswissen simulieren

Diskurswissen ist Wissen, das im Laufe eines Dialogs erworben wird. Systeme, die Diskurswissen simulieren, erscheinen dem Anrufer als hätten Sie „ein Gedächtnis“ und wirken dadurch intelligenter bzw. menschlicher.

Ein Beispiel: Für eine gewählte Aufgabe wird vom Nutzer ein Login benötigt. Will der Nutzer nach Erledigung dieser Aufgabe eine weitere Aufgabe durchführen, für die ebenfalls ein Login benötigt wird, sollte das System nicht noch einmal nachfragen, sondern „im Gedächtnis behalten“, dass der Login bereits durchgeführt wurde.

Diskurswissen findet nicht nur auf dieser Makro-Ebene des Dialogs statt. Auch im Detail findet es sich auf allen Ebenen der Sprache und trägt hier zu einem intelligent erscheinenden, flüssigen Dialog bei. Ein sehr einfaches Beispiel sind die so genannten Diskursmarker:

System	Ihre Kundennummer, bitte.
Anrufer	1376452.
System	<b>Vielen Dank. Und nun</b> bitte Ihre PIN.

Ohne Diskursmarker (z.B.: „Ihre PIN, bitte“) wird kein Bezug zu den vorangegangenen Dialogabschnitten hergestellt. Neben diesem sehr einfachen Mitteln gibt es viele weitere Möglichkeiten, Diskurswissen zu simulieren.

## Gesprochene Sprache verwenden

Geschriebene Sprache wird vollkommen anders verarbeitet als gesprochene: Der Leser verarbeitet die Sätze nicht sequentiell, sondern hat etwa die Möglichkeit, schon früh das Verb des Satzes, den wichtigsten Bestandteil, zu erfassen. Außerdem kann ein geschriebener Text beliebig oft und in beliebigem Tempo gelesen werden.

Gesprochene Sprache ist sequentiell, sie verschwindet in dem Augenblick, in dem sie gesprochen wird. Deswegen muss sie leicht verständlich aufbereitet sein: Verben beispielsweise müssen möglichst früh im Satz erscheinen und komplexe Syntax-Strukturen sind zu vermeiden.

Aber nicht nur für das Hörverstehen, sondern auch für das Hear&Feel ist der Gebrauch gesprochener Sprachstrukturen angebracht. Schriftsprachliche Floskeln (z.B. „Bitte geben Sie Ihre Kontonummer ein.“ anstatt „Ihre Kontonummer bitte.“) sind nicht nur schwerer verständlich, sondern wirken auch steif und technisch. Das System erscheint dadurch sehr unflexibel und somit weniger menschlich.

## Natürliche Äußerungen zulassen

Natürlich kann ein Nutzer mit einem System nicht so frei sprechen wie mit einem Menschen. Seine Äußerungen müssen berechenbar bleiben - dafür zu sorgen ist Aufgabe des Dialog-Designs.

Aber es gibt geschickte und weniger geschickte Wege den Nutzer einzuschränken. Hier ein weniger geschickter Weg:

System	Wenn Sie eine Überweisung tätigen möchten, sagen Sie „Überweisung“. Wenn Sie Ihre Karte sperren möchten, sagen Sie „Karte sperren“ und wenn Sie Ihre Umsätze abrufen möchten, sagen Sie „Umsätze“.
Anrufer	Karte sperren.
System	Möchten Sie Ihre Karte sperren? Sagen bitte „ja“ oder „nein“.

Das explizite Nennen der Keywords eignet sich eher dann, wenn bereits Erkennungsfehler aufgetreten sind. Ansonsten ist es besser, Keywords implizit zu formulieren. Bei ja/nein-Dialogen kann, außer im Fehlerfall, gänzlich auf das Nennen der Keywords verzichtet werden.

Der geschicktere Weg:

System	Möchten Sie eine Überweisung durchführen, Ihre Karte sperren oder Ihre Umsätze abrufen?
Anrufer	Meine Karte sperren.
System	Sie möchten Ihre Karte sperren - ist das richtig?

Der Dialog wirkt menschlicher und ist zudem wesentlich kürzer. Die Eingaben des Nutzers bleiben berechenbar und trotzdem wird das Gefühl vermittelt, frei sprechen zu können. Allerdings ist das implizite Nennen der Keywords nicht unproblematisch. Der Nutzer muss beim ersten Anhören erkennen, was genau seine Optionen sind und wie er diese auswählen kann. Hier muss neben der Textgestaltung auch die richtige Sprechweise zum Einsatz kommen (Pausen und Betonung).

### Randomized Prompting

Randomized Prompting bedeutet, dass an einer bestimmten Stelle des Dialogs nicht immer der gleiche Prompt bzw. Text abgespielt wird (z.B. „Vielen Dank. Und nun bitte Ihre PIN“), sondern zufällig ausgewählte Variationen, zum Beispiel:

- „Danke, und nun Ihre PIN, bitte.“
- „Gut - und Ihre PIN?“
- „Dankeschön. Dann brauche ich noch Ihre PIN, bitte!“

Ein einfaches Mittel, Menschlichkeit zu simulieren. Denn ein Mensch wählt in der gleichen oder ähnlichen Gesprächssituation auch nicht immer die gleichen Worte und vor allem nicht in exakt derselben Betonung.

In einigen Nutzungskontexten sollte Randomized Prompting aber gemieden werden. Wird eine Anwendung sehr häufig von einer geringen Anzahl Nutzer verwendet - zum Beispiel von Außendienstmitarbeitern eines Unternehmens zur mobilen Eingabe von Daten - ist es besser, immer den gleichen Text zu verwenden. Der Nutzer erkennt so schon nach den ersten Silben, wo er sich im Dialog befindet und kann das System mit seiner Eingabe unterbrechen, ohne den Text bis zum Ende anhören zu müssen.

### Geringe emotionale Dynamik

Sprachdialogsysteme stehen weit hinter den Fähigkeiten des Menschen zurück. Deswegen ist auch der Einsatz simulierter Emotionen problematisch: Emotionen werden als die wesentliche Eigenschaft angesehen, die Menschen von Maschinen unterscheidet.

Stimmen aber transportieren immer den Eindruck einer Emotion, so wie sie auch immer den Eindruck von Menschlichkeit erwecken. Deswegen müssen Emotionen auch ihren Platz im Dialog-Design finden. Allerdings sollten, sofern es sich nicht um eine Anwendung mit bewusst eingesetzten Humor handelt, extreme Emotionen sowie große Emotions-Sprünge gemieden werden:

System	(freundlich-offener Ton) Worum geht es? DSL-Status, Fragen zur Rechnung oder Beschwerde.
Anrufer	Beschwerde.
System	(peinlich berührter Ton) Oh, eine Beschwerde. Kleinen Moment, ich verbinde...

Hier fühlt sich der Anrufer vermutlich etwas auf den Arm genommen. Um es mithilfe des oben beschriebenen Vier-Seiten-Modells auszudrücken: Die Symptomseite und die Beziehungsseite der einzelnen Nachrichten sollte eher zurückgeschraubt werden.

### Konsistenz im Stil

Wenn man Menschlichkeit im Dialog-Design einsetzt, muss sie möglichst konsistent sein. Probleme bergen hier potentiell diejenigen Dialogpassagen, in denen die zugrunde liegende Sprachtechnologie Unnatürlichkeit hervorrufen kann:

System	Hey, hallo! Willkommen beim Klingelton-download. Gibst Du mir Deine Mitgliedsnummer?
Anrufer	<i>Na klar - das ist die 4654 343.</i>
System	Entschuldigung, das habe ich nicht verstanden. Bitte nenne Deine Mitgliedsnummer Ziffer für Ziffer.

Wie man sieht, verleitet dieser sehr menschliche Stil den Anrufer dazu, zu frei zu sprechen. Es resultiert eine Fehlerkennung und eine unnatürliche Gesprächssituation, die wegen der sehr natürlichen Begrüßung als umso unnatürlicher wahrgenommen wird. Ähnlich wie bei den Emotionen gilt auch hier, allzu große Sprünge zu vermeiden.

## Fazit

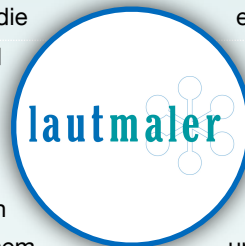
Menschlichkeit in Sprachdialogsystemen ist wünschenswert, denn richtig eingesetzt steigert sie Nutzerfreundlichkeit und Hear&Feel. Falsch eingesetzt jedoch, wirkt sie dem entgegen: Der Nutzer kann nicht mehr einschätzen, welche Äußerungen das System versteht und welche nicht und zugleich wirkt das System in seiner Simulation unglaubwürdig.

Wenn aber die richtigen menschlichen Elemente richtig eingesetzt und aufeinander abgestimmt werden, wird das Optimum erreicht und der Dialog gewinnt erheblich an Natürlichkeit und Intuitivität.



## Die Lautmaler - eine Agentur für Voice User Interface Design

Die Lautmaler GbR ist eine in Berlin ansässige Agentur für Voice User Interface Design. Kerngebiet ist die Gestaltung der Schnittstelle zwischen Nutzer und System. Dazu gehört neben dem Dialog- und Audio-Design auch der Bereich Speech Science, der Leistungen wie Grammatik-Design, linguistische Datenredaktion, phonetische Transkription sowie Erkennen- und TTS-Tuning umfasst. Mit diesem ganzheitlichen Ansatz finden die Lautmaler für jedes



System den Königsweg zwischen intuitiver Bedienbarkeit einerseits und technisch Möglichem andererseits. Zum Team der Lautmaler gehören die Autoren Manja Baudis und Zeno Wolze. Manja Baudis ist Computerlinguistin und Germanistin und bei den Lautmalern für Speech Science und Marketing zuständig. Zeno Wolze ist Linguist und Phonetiker und bei den Lautmalern zuständig für Dialog-Design und Sales.

### Kontakt

office: +49 (0)3049 85 57 10  
email: kontakt@die-lautmaler.de  
web: www.die-lautmaler.de

© Die Lautmaler 2008 - Alle Rechte vorbehalten

#### Manja Baudis

office: +49 (0)30 63 21 80 40  
mobil: +49 (0)171 6 54 92 30  
email: manja.baudis@die-lautmaler.de

#### Zeno Wolze

office: +49 (0)30 49 85 57 10  
mobil: +49 (0)177 6 91 15 44  
email: zeno.wolze@die-lautmaler.de

### DIE LAUTMALER

Florastraße 78  
13187 Berlin