

**Vortragsbericht zu »Telefonieren mit einem Computer: Telefonbasierte natürlichsprachliche Dialogsysteme«
gehalten von Evelyn Thar (Universität Zürich) am 01.12.2009 im Rahmen von LinguA an der Leibniz Universität Hannover**

Für die Einen sind sie ein Segen, für die Anderen ein Fluch: „Sprachcomputer“ am Ende der Kundenhotline sparen dem Dienstleister Personalkosten und garantieren (nur vermeintlich?) eine zügige und effiziente Bearbeitung der meisten Bedürfnisse ihrer Kunden. Für den Anrufer am anderen Ende der Leitung sind sie nur zu häufig Anlass zum Verdross – wer lässt sich schon gerne zum bloßen Erzeuger eines standardisierten Inputs herabwürdigen: Das vom Anrufer als individuell und speziell empfundene Anliegen lässt sich offenbar in häufig wiederkehrende und einfach klassifizierbare Kategorien fassen, weshalb im Zuge des Telefongesprächs wenig bis keinen Raum für Wenns und Abers, für unerwartete Abweichungen von den in Form von Auswahlmenüs dargebotenen Routineabläufen bleibt. Hatte man es als Kunde in der Frühzeit solcher Dialogsysteme dann auch noch mit eindeutig elektronisch generierten Stimmen zu tun, sank die Kooperationsbereitschaft vieler Kommunikationspartner der Spezies *homo sapiens* im Verlauf des Telefongesprächs auf Null. Auf so viel Widerborstigkeit ihrer menschlichen Kunden und der damit verbundenen potentiellen Gefährdung ihrer Umsätze mussten die Anbieter nun reagieren.

Dies geschah, indem die elektronischen Kommunikationspartner ein immer menschlicheres Antlitz erhielten, beispielsweise mithilfe einer stärkeren Annäherung ihrer Sprachproduktion an den Klang der menschlichen Stimme durch die Aufzeichnung der Systemansagen (*Prompts*) durch professionelle (menschliche) Sprecher/innen, aber auch durch Programmierung immer flexiblerer und am menschlichen Kommunikationsverhalten orientierter Formulierungen bei Berücksichtigung einer größtmöglichen Nähe zu zwischenmenschlichen Gesprächsabläufen. Auf technischer Ebene werden diese Routinen telefonischer Sprachdialogsysteme realisiert durch interaktive Spracherkennungsprogramme, welche entweder auf Tastatureingaben oder auf akustischen Input reagieren und die Eingabe bzw. Ansage des menschlichen Anrufers entsprechend des Inhalts seines gegebenen Signals vorqualifizieren. Innerhalb des Baums möglicher Optionen wird die Anfrage daraufhin in den mutmaßlich richtigen Zweig weitergeleitet. Im Hintergrund werden diese Operationen ermöglicht durch Programmierung nicht nur umfangreicher Lexika, sondern auch komplexer Regelwerke, die grammatische Strukturen gleichermaßen berücksichtigen wie semantische Netze und Einheiten auf pragmatischer bzw. diskursanalytischer Ebene. Auf der Benutzeroberfläche hingegen sieht man sich möglichst realistisch und menschlich konstruierten Gesprächspartnern gegenüber, welcher nicht nur auf akustischer Ebene so menschlich wie möglich klingt, sondern auch das Personalpronomen „ich“ verwendet und in der Lage ist, typische para-verbale Rückmeldungssignale der zwischenmenschlichen Kommunikation wie „Hmm“, „Oh“, oder „M-mmh“ zu geben. Gerne wird dieser elektronische Gegenüber auch mit Namen, Bild und (fiktivem) Lebenslauf ausgestattet, um ihm (oder häufiger: ihr) ein so menschliches Gesicht wie nur möglich zu geben. Beispiele hierfür sind die Schweizer Datingportal im Internet „Lucy“ oder auch die virtuelle Studienberaterin des Deutschen Akademischen Austausch Dienstes (DAAD), „Luzie“. Der Vorteil dieser fingiert menschlichen Gesprächspartner liegt auf der Hand: Der Benutzer tritt entspannter in die Kommunikation ein, spricht frei und ohne Einschränkung mit dem Dialogsystem, was durch intuitives Design der zugrundeliegenden Spracherkennungssoftware ermöglicht wird.

Leider ist diese immer noch weit davon entfernt, perfekt in der Imitation menschlichen Sprachverständnisses und Gesprächsverhaltens zu sein. Im Unterschied zur medial schriftlichen Online-Kommunikation, in der die Eingaben durch den menschlichen Benutzer meist mit Hilfe der Tastatur oder durch Mausclicks erfolgen, müssen bei der Interaktion mit telefonbasierten Dialogsystemen die Spezifika des Telefongesprächs als institutionalisiertem Mediengesprächstyps berücksichtigt werden, und zwar nicht nur hinsichtlich des technischen Übertragungskanal, sondern auch der besonderen Charakteristika und Strategien dieses Diskurstyps. Da es sich um einen Gesprächstyp handelt, der sich eines ausschließlich akustischen Kommunikationskanals bedient, welcher zwischen zwei separaten Wahrnehmungs- und Handlungsräumen

vermittelt, kommt den nonverbal-vokalen Äußerungskomponenten wie Tonhöhe, Lautstärke und Betonung eine größere Bedeutung zu. Schweigephasen sind hierbei deutlich problematischer als im Falle von *face-to-face*-Kommunikation. Für die telefonbasierte Kommunikation mit maschinellen Dialogsystemen stellt sich dem Anbieter dieses Mediums zudem die Anforderung, die Bediensprache so natürlich wie möglich zu gestalten und zudem der in der Regel eingeschränkten Fachkompetenz des Kunden insofern Rechnung zu tragen, als man ihn mit möglichst wenig domänenspezifischem Fachvokabular konfrontieren sollte. Problematisch bei der Interaktion mit Maschinen ist ferner, dass das Dialogsystem bei der „Formulierung“ seiner Äußerung meist auf wenig Kontext- und Weltwissen zurückgreifen kann: Es verfügt nur über so viel Weltwissen, wie ihm von vorneherein einprogrammiert wurde und kann kontextuelle Informationen nur insoweit reinferieren, wie es – auf Basis komplexer Rechenoperationen – in der Lage ist, sie den Äußerungen seines menschlichen Gesprächspartners zu entnehmen. Die Teilnahme an Aushandlungsprozessen einer lokal gültigen Gesprächsordnung ist der Maschine nicht möglich, ebenso wenig wie die Inferenz und Berücksichtigung von nicht verbal (oder ggf. zusätzlich visuell) zur Verfügung gestellter Informationen zum situativen Kontext.

Der gesamte Kommunikationsrahmen in der Mensch-Maschine-Interaktion ist aufgrund all dieser Probleme nur beschränkt funktionsfähig; Verbesserungen des Systemdesigns müssen daher gerade an diesen Punkten ansetzen, um den Erwartungen der Benutzer in stärkerem Maße gerecht zu werden und infolgedessen Gesprächsabbrüche frustrierter Anrufer zu vermeiden. Dieser Forderung liegen die Thesen zugrunde, dass sich

1. die Erwartungen der Benutzer (Anrufer) auf ihre vorhandene Kompetenz auf dem Gebiet der Telefonkommunikation als spezifischem Mediengesprächstyp stützen und
2. keine Aussagen über ein generelles zu erwartbares Benutzerverhalten gemacht werden können, da unterschiedliche Kommunikationsstrategien angewendet werden und Kommunikationsstrategien nicht konsequent verfolgt sondern verworfen werden, wenn sie nicht erfolgreich sind¹

Das bedeutet in der Praxis, dass in den Ansagen des Dialogsystems die Turn-Grenzen (also die Grenzen der jeweiligen Redebeiträge der Gesprächsteilnehmer) stark genug markiert sein müssen, damit der Benutzer weiß, wann er sprechen darf und wann nicht. Ferner muss bei der Konstruktion des Dialogsystems die Überlegung einfließen, ob die Unterbrechung von Systemansagen gestattet werden soll oder nicht, da Unterbrechungen von menschlichen Kommunikationsteilnehmern in der Regel als aggressiv und unangenehm empfunden werden, selbst wenn nur einer Maschine ins Wort gefallen wird. Empirische Tests zeigen, dass die Regeln zwischenmenschlichen Gesprächsverhaltens offenbar seitens der menschlichen Kommunikanten auch an die Kommunikation mit Maschinen angelegt werden, selbst wenn ihnen bewusst ist, dass es sich bei dem Gegenüber um eine Maschine handelt. Auch werden Rückmeldesignale wie „ja“, „mmh“, „danke“ und „das ist lieb“ seitens des Benutzers verwendet, welche als ungeplante Eingabe zu mehrstufigen Fehlereskalationen beim Dialogsystem führen können. Ähnliche Folgen haben die bereits erwähnten längeren Schweigephasen auf Seiten des Benutzers sowie Eröffnungs- und Beendigungssequenzen des Gesprächstyps „Telefongespräch“. „Ungeplante“ Verhaltensweisen des Anrufers wie Rückmeldungssignale oder Eröffnungs-/Beendigungssequenzen können jedoch nicht unterdrückt werden und sollten daher als bereits bekannte Kommunikationsmuster, die im Input möglicherweise auftreten können, behandelt und als zusätzliche, gesprächsanalytisch basierte Komponenten beim Design des Dialogsystems berücksichtigt werden (und sei es nur durch Ignorieren von Unterbrechungen durch Deaktivierung der sog. „Barge-In-Möglichkeit“, um das Einsetzen einer Fehlereskalation zu vermeiden). Ferner kann dem Benutzer die Interaktion durch klare, präzise Ansagen seitens des Systems erleichtert werden, welche es dem Anrufer ermöglicht, das System als *Instrument* zu benutzen, anstatt ihn dazu zu verleiten, in dem Dialogsystem einen quasi menschlichen Kommunikationspartner zu sehen, welcher über entsprechend flexible Gesprächsstrategien verfügt. Dabei sollten dem menschlichen Benutzer klare

¹ vgl. auch Fischer, K. (2006): *What Computer Talk Is and Isn't*. Saarbrücken, AQ-Verlag.

Hilfestellungen ebenso zur Bedienung des Systems gegeben werden wie Signale zur Art und Zeitpunkt der nächsten Äußerung des Benutzers.

Fazit: Die bislang angestrebte größtmögliche Nähe maschineller Dialogsysteme zu menschlichen Sprechern scheint nicht unter allen Gesichtspunkten von Vorteil zu sein. Für die erfolgreiche Kommunikation mit natürlichsprachlichen Dialogsystemen vor allem im Bereich Telefonkommunikation ist es von großer Bedeutung, dass sein Charakter als reines Hilfsmittel zur Erlangung externer Kommunikationsziele im Vordergrund steht. Als solches muss es dem natürlichen Gesprächsverhalten der menschlichen Kommunikanten mit all seinen Diskursmerkmalen und -strategien maximal Rechnung tragen, was durch die Implementierung gesprächsanalytischer Komponenten in das Dialogsystem erreicht werden könnte.

Alexa Mathias