

Der Ton macht den Einkauf zum Erlebnis

Auf die Emotion in der Stimme des Sprachassistenten kommt es beim Einkauf an

SKommunikation zwischen Menschen und Computer findet immer häufiger über Stimme statt. Aus der Forschung ist bekannt, dass die Emotion der Stimme den Verkaufsprozess entscheidend beeinflussen kann. Doch gilt das auch für die synthetische Stimme von Voice Assistants? Dr. Carolin Kaiser und René Schallner vom Nürnberg Institut für Marktentscheidungen (NIM) wollten dies herausfinden und stellen fest: Auf die Emotion in der Stimme des Sprachassistenten kommt es an. Eine fröhliche Stimme schafft Vertrauen, eine begeisterte Stimme führt zu Impulskäufen auch teurer Produkte. Eine emotionale Computerstimme erzielt dabei eine vergleichbare Wirkung wie eine menschliche Stimme.

Sprachassistenten boomen: Nach Schätzungen von Marktforschern wird sich die Zahl der Geräte in Haushalten weltweit bis zum Jahr 2023 verdreifachen. Allein im vierten Quartal 2019 gingen weltweit rund 56 Millionen Stück der intelligenten Lautsprecher von Amazon, Google und anderen Anbietern über den Ladentisch. Und dies, obwohl die aktuelle Generation der Sprachassistenten immer noch mit einer wenig emotionalen und eher gleichgültigen Stimme spricht. Dabei belegt eine aktuelle Studie des NIM, dass sich 60 Prozent der intensiven Nutzer mehr Emotionalität, Fröhlichkeit oder Abwechslung von ihrem Sprachassistenten wünschen. Und die Hersteller von



ABB: ANDRUS / FOTOLIA

Sprachassistenten handeln bereits. Aktuelle Patentanmeldungen sowie Ankündigungen von marktführenden Smartspeaker-Anbietern zeigen, wohin der Weg von Alexa & Co führt: Mittels künstlicher Intelligenz wird deren Stimme schon in naher Zukunft menschlicher und emotionaler werden.

Zwar ist noch nicht klar absehbar, wie sehr die Geräte als Verkaufskanal das Machtgefüge zwischen Konsument, Händler und Hersteller verändern werden, aber mit Emotionalität in der Stimme sind Sprachassistenten im Kernbereich von Kaufentscheidungen angekommen. Denn schon lange ist umfassend belegt, dass in der Mensch-Mensch-Interaktion Emotionen – beispielsweise ausgelöst durch Stimme – ansteckend sind und das Kaufverhalten beeinflussen. So neigen beispielsweise emotional erregte Kunden dazu, Kaufentscheidungen impulsiver zu treffen. Glückliche Kunden wiederum bewerten ihre Zufriedenheit mit Produkten deutlich besser. Für das NIM mit seinem Forschungsfokus auf dem Thema Marktentscheidungen lautet daher die Frage: Beeinflusst der fröhliche oder gar begeisterte Tonfall eines Sprachassistenten das Kaufverhalten eines Konsumenten? Die hier vorgestellte Studie gibt darauf Antworten und liefert für das Marketing relevante Erkenntnisse.

Um belastbare Daten für dieses noch wenig erforschte Gebiet zu erhalten, war ein mehrstufiges Vorgehen notwendig, das sich an dem wissenschaftlich etablierten Stimulus-Organismus-Reaktions-Modell orientierte. Als Stimulus diente eine emotionale Stimme des Sprachassistenten. Gemessen wurden die Emotionen der Versuchsteilnehmer in den Dimensionen Valenz (positiv vs. negativ) und Arousal (ruhig vs. emotional erregt). Anschließend wurde die konkrete Reaktion auf den Stimulus mittels Beobachtung von Kaufverhalten und der Abfrage von Kaufeinstellungen ermittelt.

Voice-Assistenten lernen Emotionen

Für den ersten Schritt der Studie – die Erstellung von emotionaler Tonalität für Sprachassistenten – mussten umfassende Modelle entwickelt und getestet werden, da noch keine emotional sprechenden Sprachassistenten auf dem Markt sind. Mithilfe neuester technischer Verfahren des Deep Learnings entstand ein Speech-Synthesizer, der beliebigen Text in drei emotionalen Tonalitäten sprechen kann: fröh-

Die Autoren



Dr. Carolin Kaiser ist Senior Researcher in der Data Science und Behavioral Science Group am Nürnberg Institut für Marktentscheidungen (NIM). Zuvor war sie als wissenschaftliche Mitarbeiterin am Institut für Wirtschaftsinformatik der Universität Erlangen-Nürnberg tätig, wo sie im Bereich der automatisierten Meinungsanalyse im Social Web promovierte.

carolin.kaiser@nim.org



René Schallner ist Senior Researcher und Leiter des Tech-Labs beim Nürnberg Institut für Marktentscheidungen (NIM). Zuvor sammelte er zwei Jahrzehnte lang tiefes Know-how als international aktiver, selbstständiger Software-Entwickler im Telekom- und Medizintechnik-Bereich.

rene.schallner@nim.org

lich, begeistert und unbeteiligt. Im Vergleich zu anderen Verfahren sind nur wenige Ressourcen nötig, um die generierte Emotion in der Stimme sehr natürlich klingen zu lassen: Etwa eine Stunde Sprachaufnahmen sowie zwölf Stunden Rechenzeit.

Zur Evaluierung des Speech-Synthesizers wurden in jeder der drei Emotionen zehn Sätze erstellt, die für Einkaufssituationen typisch sind. Die Überprüfung der synthetischen Sprachbeispiele durch Versuchsteilnehmer ergab bereits ein eindeutiges Ergebnis: 224 Teilnehmer auf einer Crowdsourcing-Plattform erkannten die entsprechende Tonalität der synthetischen Stimme von 71 Prozent der Sätze richtig. Sprachkenntnisse waren nicht gefordert. Die 121 deutschsprachigen Teilnehmer der Universität Hamburg erkannten die Tona-

lität sogar in 90 Prozent der Sätze korrekt. Das Ergebnis, dass sich die synthetisch hergestellten Stimmen einzelnen Emotionen zuordnen lassen, bildete den Ausgangspunkt für den nächsten Schritt der Forschungsstudie: Jetzt wurden die Auswirkungen der Tonalität der menschlichen Stimme sowie der Tonalität des Sprachassistenten auf das Einkaufsverhalten von Studienteilnehmern untersucht. Dies erfolgte mit einem eigens programmierten Online-Shop, der – entweder mit oder ohne Sprachunterstützung – die zentralen Aspekte des Kaufprozesses abdeckte, darunter Auswahlmöglichkeiten, Warenkorb und Bezahlung.

Für die Studie wurde ein Online-Shop als Android-Anwendung programmiert, der eine Alexa-ähnliche Sprachausgabe in den Verkaufsprozess integrierte. Die insgesamt 213 Teilnehmer bekamen ein Budget von jeweils zehn Euro und sollten insgesamt drei unterschiedliche Büroartikel einkaufen. Die Produkte, Radierer, Stift, Notizblock, wurden in verschiedenen Qualitäten mit unterschiedlichen Preisen angeboten. In jeder Produktkategorie wurde ein Premiumprodukt mit Zusatzfunktion zu einem erhöhten Preis angeboten. Die Budgetverwendung über die drei Produkte hinaus war vollkommen freigestellt und ermöglichte so freie und spontane Kaufentscheidungen. Zum Abschluss des Kaufs wurde zusätzlich ein Elektroartikel als Sonderangebot offeriert, den die Teilnehmer optional kaufen konnten und der auf potenzielle Impulskäufe abzielte. Um die Kaufentscheidungen möglichst real zu gestalten, bekamen die Teilnehmer am Ende des Experiments alle gekauften Produkte ausgehändigt und ihr Restbudget in bar ausgezahlt.

Die Teilnehmer erhielten vor und nach dem Einkaufsprozess einen Fragebogen, mit dem die emotionale Valenz (positiv vs. negativ) abgefragt wurde. Während des

Welche Wirkung hat der emotionale Tonfall eines Sprachassistenten?

Stimulus-Organismus-Reaktions-Modell



Quelle: NIM

planung&analyse 2/2020

Einkaufs wurde die Herzfrequenz als Indikator für das emotionale Arousal (ruhig vs. erregt) gemessen. Nach Abschluss des Bestellvorgangs erfolgte eine weitere Befragung, die die Nutzererfahrung, das Vertrauen und die Zufriedenheit erfasste. Insgesamt konnten 206 Datensätze für die Studie verwertet werden.



Sprache führt zu schnelleren Kaufentscheidungen

Eine erste relevante Erkenntnis über Kaufentscheidungen betrifft das Einkaufsverhalten. Seit Jahren kaufen Konsumenten online ein und haben sich dabei ein bestimmtes Vorgehen angewöhnt. Herkömmliche Online-Shops sind jedoch rein textbasiert. Wie verändern sich nun Kaufeinstellung und Kaufverhalten, wenn plötzlich Sprachassistenten mit dem Konsumenten beim Einkauf sprechen?

Die NIM-Studie zeigt: Voice Shopping – egal ob mit aufgezeichneter menschlicher Stimme oder mit synthetischer Stimme und egal mit welcher der drei Emotionen – beschleunigt den gesamten Einkaufsprozess und fördert impulsives Kaufverhalten im Vergleich zu herkömmlichem Shopping ohne Sprachunterstützung. Die Daten belegen: Konsumenten treffen schneller Kaufentscheidungen und sind eher geneigt – einem plötzlichen Drang folgend – spontan Produkte zu kaufen, die sie eigentlich nicht vorhatten zu kaufen. Bezüglich der Einkaufseinstellung gibt es jedoch keine wesentlichen Unterschiede zwischen Voice Shopping und traditionellem textbasierten Shopping: Das Vertrauen zum Shop, die User Experience und die Kundenzufriedenheit sind vergleichbar.

Ein weiteres Studienergebnis betrifft die Wahrnehmung von Stimme. Zwar empfinden die Studienteilnehmer die aufgenommene menschliche Sprache als natürlicher und demzufolge auch als sympathischer. Aber die emotionale Wahrnehmung zwischen menschlicher Stimme und synthetischer Stimme eines Sprachassistenten unterscheidet sich nicht signifikant. Und bei den Auswirkungen auf das Einkaufsverhalten und die Kaufeinstellung gibt es zwischen einer emotionalen synthetischen Stimme und einer aufgezeichneten menschlichen Stimme keine wesentlichen Unterschiede. Für das Marketing von Shopbetreibern bedeutet dies, dass eine emotionale synthetisierte Stimme eine gute und preiswerte Alternative zu aufwändig eingesprochenen Textbausteinen sein kann.



Der Tonfall beeinflusst das Kaufverhalten

Beim konkreten Einkaufsverhalten bestätigt die Studie das aus der Interaktion von Mensch zu Mensch vielfach dokumentierte Modell. Emotionen übertragen sich und beeinflussen Kaufverhalten und Kaufeinstellung. Dies gilt nach den Ergebnissen der vorliegenden Studie auch von Computer zu Mensch. Eine fröhliche Stimme steigert bei den Studienteilnehmern die emotionale Valenz am meisten, gefolgt von einer begeisterten Stimme. Auf Platz drei folgt die unbeteiligte Stimme. Das durch eine fröhliche Stimme ausgelöste positive Gefühl führt dazu, dass die Teilnehmer die Benutzererfahrung und

So klingen die NIM-Stimmen

Cineasten mögen sich angesichts dieser Möglichkeiten von künstlicher Stimme an HAL 9000 erinnern, den sehenden und sprechenden Großrechner aus dem Film von Stanley Kubrick „2001: Odyssee im Weltraum“. Auch HAL 9000 – erschienen 1968 – nutzt unterschiedliche Tonalitäten seiner Stimme, um Emotionen auszudrücken. Allerdings endet für HAL 9000 seine Macht, als ihm der Mensch Schritt für Schritt seine Rechenleistung abschaltet, da er seine menschlichen Kollegen ermordet hatte. Eine Vision, die Sprachassistenten angesichts ihrer millionenfachen Verbreitung und vielfach entgegengebrachter Zuneigung kaum widerfahren dürfte. Die Aussagen von HAL wurden damals allerdings von einem Schauspieler eingesprochen und so verändert, wie sich Hollywood vorstellte, dass künstliche Sprache einmal klingen wird. Dank modernster Deep-Learning-Verfahren klingt künstliche Stimme jedoch heute viel natürlicher als damals prognostiziert.



Interaktion am besten einschätzen, dem Shop in höchstem Maße vertrauen und am zufriedensten mit Shop und den gekauften Produkten sind.

Wer im Marketing jedoch den Verkaufserfolg steigern will, sollte eine begeisterte Stimme einsetzen. Bei dieser Tonalität steigt der Puls während des gesamten Einkaufs – und auch bei Erhalt des Sonderangebots – am stärksten an. Das Arousal ist signifikant höher als bei der fröhlichen oder unbeteiligten Stimme. Die durch die begeisterte Stimme ausgelöste emotionale Erregung führt dazu, dass die Teilnehmer deutlich impulsiver das Sonderangebot kaufen, insgesamt mehr Geld ausgeben und zudem mehr Premiumprodukte kaufen.

Ein fröhlicher oder gar begeisterter Tonfall eines Sprachassistenten beeinflusst das Kaufverhalten und die Kaufeinstellung eines Konsumenten also positiv: Das Stimulus-Organismus-Response-Modell, das in der Mensch-Mensch-Interaktion als Standardmodell gelten kann, behält auch in der Mensch-Computer-Interaktion seine Gültigkeit. Emotionsansteckung funktioniert also nicht nur zwischen Menschen, sondern auch zwischen Computern und Menschen. Oder anders formuliert: Klingen Alexa, Siri und Co fröhlich oder begeistert, haben sie Auswirkungen auf die Emotionen der Konsumenten. Diese Emotion der Konsumenten zeigt sich in deren Valenz und Arousal, was sich wiederum auf ihre Kaufeinstellung und das Kaufverhalten auswirkt.

Marketingmanager sollten daher die weitere Entwicklung der Sprachassistenten aufmerksam verfolgen und potenzielle Einsatzmöglichkeiten evaluieren. Denn dass ein Smart Speaker inzwischen durchaus die Rolle des Alltagsbegleiters übernehmen kann, hat eine weitere NIM-Studie ermittelt: Die Nutzer wünschen sich hier allerdings, dass Alexa und Co mit mehr Emotion in der Stimme auf sie eingehen und der Datenschutz generell bei diesen Geräten gewährleistet ist.

Weitere Forschungsarbeiten mit größerer Datenbasis sind noch erforderlich, um die Ergebnisse statistisch weiter zu validieren. Deshalb setzt das NIM die Forschung in diesem Themenbereich fort: Bereits in der konkreten Vorbereitung befindet sich eine Verbreiterung des Befragungssample. Außerdem ist eine Ausweitung auf einen internationalen Kontext geplant. Erforscht werden soll, ob kulturelle Unterschiede bei einem gleichen emotionalen Tonfall des Sprachassistenten zu anderen Marktentscheidungen führen. ■