

Affinitätsdiagramm – Qualitative Auswertung einer FAS-Nutzerstudie

Christoph Klöffel, Christian Purucker, Frederik Naujoks, Ann-Kathrin Kraft

Würzburger Institut für Verkehrswissenschaften (WIVW GmbH, Veitshöchheim)

Zusammenfassung

Bei der Durchführung von empirischen Nutzerstudien fallen häufig große Mengen qualitativer Daten an, etwa wenn Benutzungstests mit der „Thinking-Aloud“-Methode durchgeführt werden, bei welcher Probanden ihre Bedienhandlungen kommentieren. In der vorliegenden Arbeit wird die Anwendung der Methode des Affinitätsdiagramms am Beispiel eines Nutzertests eines Fahrerassistenzsystems (FAS) dargestellt. Im Rahmen der Studie wurden 590 eigenständige Aussagen erfasst, welche die Eindrücke während der Fahrt mit dem Fahrerassistenzsystem im realen Anwendungsfall widerspiegeln. Diese Aussagen werden mit Hilfe eines Affinitätsdiagramms in die wichtigsten Bedürfnisse der Nutzer kategorisiert und dargestellt. Die bottom-up Kategorisierung mittels eines Affinitätsdiagramms wird mit anderen Klassifikationsmethoden verglichen, welche einen top-down Ansatz verfolgen. Es zeigt sich, dass die Methode des Affinitätsdiagramms gegenüber den top-down Methoden am vorliegenden Datenmaterial zu einer deutlicheren Gleichverteilung der Aussagen über die generierten Kategorien führt als die Einstufung der Aussagen in vorgegebene Kategorien. Zudem können einzelne Themenbereiche detaillierter aufgelöst und betrachtet werden, auch weil die Ebenentiefe bei der Auswertung flexibel angepasst werden kann. Die Methodik ist darüber hinaus auch auf qualitatives Datenmaterial anderer Themenbereiche anwendbar.

1 Einleitung

Bei der Durchführung von empirischen Nutzerstudien fallen mitunter große Mengen unstrukturierter qualitativer Daten an, etwa im Rahmen von offenen Interviews, freien Aussagenprotokollen und ungerichteten Beobachtungen von Nutzerhandlungen. Dieses zunächst ungeordnete qualitative Datenmaterial muss in der Regel für weiterführende Auswertungen quantifiziert werden, was beispielsweise in einem ersten Schritt über die Klassifizierung des Datenmaterials geschehen kann.

Im konkreten Fall geht es um Datenmaterial, welches im Rahmen einer Nutzerstudie mit Hilfe der „Thinking-Aloud“-Technik generiert wurde. In der Studie waren Fahrer während der Nutzung eines neuen Fahrerassistenzsystems im realen Straßenverkehr dazu aufgefordert, ihre

Veröffentlicht durch die Gesellschaft für Informatik e.V. 2016 in
B. Weyers, A. Dittmar (Hrsg.):
Mensch und Computer 2016 – Workshopbeiträge, 4. - 7. September 2016, Aachen.
Copyright © 2016 bei den Autoren.
<http://dx.doi.org/10.18420/muc2016-ws08-0003>

Gedanken in Bezug auf das System und die Systemnutzung laut zu artikulieren; diese Aussagen wurden zunächst von einem Versuchsleiter protokolliert und später der nachfolgend beschriebenen Analyse mittels des Affinitätsdiagramms unterzogen.

Die Analyse mittels eines Affinitätsdiagramms erfolgt daten- bzw. nutzeraussagengetrieben, d.h., entsprechend eines bottom-up-Prozesses. Dem gegenüber stehen top-down-Verfahren, welche zunächst aufgrund theoretischer und inhaltlicher Überlegungen im Vorfeld Klassifikationskategorien bereitstellen, in welche die Aussagen dann einsortiert werden. Diese grundsätzlichen Herangehensweisen werden in der vorliegenden Arbeit ebenfalls differenziert betrachtet und gegenübergestellt.

2 Theoretischer Hintergrund

Im Folgenden werden zwei Ansätze zur Klassifizierung qualitativen Datenmaterials vorgestellt: Ein top-down Verfahren, welches Kategorien anhand inhaltlicher und theoretischer Überlegungen bereitstellt sowie ein bottom-up Verfahren, welches einen datengetriebenen Klassifizierungsansatz verfolgt.

2.1 Theoriegeleitete top-down Kategorisierung

In der Literatur sind top-down Ansätze zur Datenkategorisierung weit verbreitet (cf. Banks & Stanton, 2015; 2016). Diese Verfahren basieren in der Regel darauf, dass ausgehend von inhaltlichen und theoretischen Überlegungen oder bestehenden Arbeiten Kategorien generiert werden, welchen die einzelnen Aussagen oder Beobachtungen dann zugewiesen werden können – das Verfahren ist zeiteffizient und die Kategorisierung durch die Vorgabe vorhersehbar und steuerbar. Auf der anderen Seite kommt es hier häufig zu einem Informationsverlust, da die Daten stark hypothesengetrieben betrachtet werden und unerwartetes Datenmaterial nur unzureichend berücksichtigt wird – der eigentliche Vorteil aus der Verwendung qualitativer Daten, die Hypothesengenerierung, wird somit nicht ausgeschöpft. Exemplarisch sei die Anwendung des Ansatzes von Banks und Stanton (2015) dargestellt, in welchen eine solche Kategorisierung der Daten in vorher festgelegte Kategorien stattfindet.

Banks und Stanton (2015) nutzten für die Auswertung von Kommentaren ein sowohl theoriegesteuertes als auch datengesteuertes Kodierschema. In ihrer Studie wurden verbale Kommentare analysiert, die während der Fahrt von $N = 2$ Probanden mit einem prototypischen Fahrerassistenzsystem aufgezeichnet wurden. Das verwendete Schema zur Datenkodierung und -kategorisierung besteht aus sieben Kategorien und basiert auf Arbeiten zur „systemic situation awareness“ Forschung. In der ursprünglichen Arbeit ist die zu Grunde liegende Datenmenge eher gering (ca. 100 Aussagen) und die einzelnen Kategorien sind nach Zuordnung der einzelnen Aussagen relativ gleichmäßig besetzt.

Bei der Anwendung dieses Verfahrens auf größere Datenmengen, auch wenn diese aus dem gleichen Themenfeld kommen und in vergleichbarem Versuchsetting erhoben wurden, zeigen

sich Schwächen in der Übertragbarkeit des top-down Kodierschemas. Wie am Beispiel der Klassifizierung der Daten der vorliegenden Studie (vgl. Kapitel 3.1) in Abbildung 1 zu sehen ist, ist es bei nur sieben vorgegebenen Kategorien möglich, dass einzelne Kategorien deutlich überrepräsentiert sind. Dadurch gehen möglicherweise Informationen verloren, da Abweichungen bzw. detailliertere Meinungen innerhalb dieser Kategorie nicht berücksichtigt werden. In einem Affinitätsdiagramm (vgl. Kapitel 3.2) entspricht diese Art der Kategorisierung der übergeordneten 3. Ebene (siehe Abbildung 2). Es fehlen hierbei weitere detaillierte Unterteilungen der Ebenen 2 und 1, welche konkrete Nutzeraussagen abbilden; beispielsweise können unmittelbar keine Aussagen getroffen werden, welche Schwierigkeiten beim Systemumgang („Functionality Issues“) aufgetreten sind.

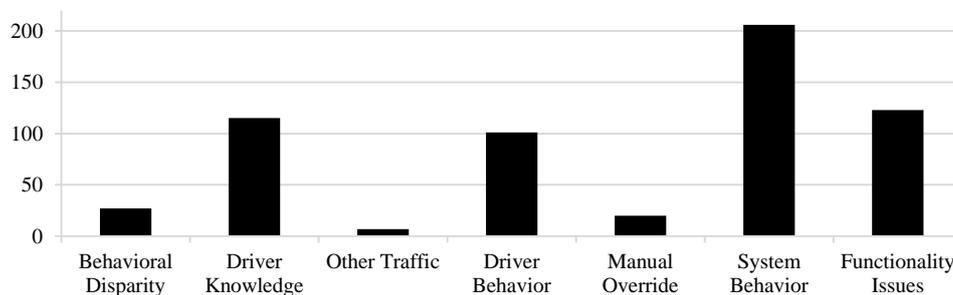


Abbildung 1. Anzahl der Aussagen nach der Klassifizierung mit dem Kodierschema nach Banks und Stanton (2015)

2.2 Bottom-up Kategorisierung mittels des Affinitätsdiagramms

Die Methode des Affinitätsdiagramms ist ein benutzerzentrierter Kategorisierungsansatz, welcher für eine große Bandbreite auch umfangreicher qualitativer Daten zum Einsatz gebracht werden kann. Das von Jiro Kawakita 1967 entwickelte Affinitätsdiagramm (Kawakita, 1967; cf. Beyer, & Holtzblatt, 1997; cf. Holtzblatt, Wendell, & Wood, 2004; cf. Kuniavsky, 2003) wird datengesteuert entwickelt und ist dabei speziell auf die Bedürfnisse des Nutzers ausgerichtet. Es wird gerne benutzt, um eine zusätzliche Einsicht in qualitative Daten zu ermöglichen und Nutzungsmuster zu erkennen (Shafe, Smith, & Linder, 2005). Regelmäßig angewendet wird das Affinitätsdiagramm innerhalb des Qualitätsmanagements zur Strukturierung qualitativer Daten (cf. Pyzdek, 2003). Auch in den automobilen Kontext hat das Verfahren bereits Einzug erhalten, etwa bei der Konzeption einer Blickbewegungsassistenz im Fahrzeug (Gärtner, Meschtscherjakov, Maurer, Wilfinger, & Tscheligi, 2014), der Analyse von ästhetischen Faktoren der Gestensteuerung eines Fahrzeuginfotainmentsystems (Wu, Tan, & Zhao, 2013) oder einem Ansatz der Gamifizierung des sicheren Fahrens, um unsichere Fahrmanöver zu reduzieren (Rodríguez, Ibarra, Roa, Curlango, Bedoya, & Montes, 2014).

Das Affinitätsdiagramm wird datengesteuert und informationsverlustfrei generiert. Aufgrund des Vorgehens bei der Gruppierung und Analyse des Datenmaterials ist diese Methode

gleichwohl zeitlich aufwändig; der Aufwand steigt hierbei exponentiell mit der Datengröße an.

3 Methode

Das Vorgehen bei der Auswertung der vorliegenden Daten mittels des Affinitätsdiagramms wird im Folgenden detailliert dargestellt.

3.1 Material

Die verwendeten Daten stammten aus einer Nutzerstudie, in der die Bedienbarkeit eines Fahrerassistenzsystems (FAS) während der Fahrt im Realverkehr untersucht wurde. Das FAS dient dazu, dem Fahrer ein teilautomatisiertes Fahrerlebnis zu vermitteln, indem vom System zeitlich begrenzt sowohl Längs- als auch Querführung übernommen wird. Aufgrund der starken zeitlichen Begrenzung der Verfügbarkeit des Systems ist dieses aber grundsätzlich noch dem Bereich der Fahrerassistenz zuzuordnen; der Fahrer muss zu jeder Zeit bereit sein, die Fahrzeugführung wieder zu übernehmen. Ziel der Studie war es, herauszufinden wie die Bedienbarkeit des Systems von den Versuchspersonen (VP) eingeschätzt werden würde. $N = 33$ Probanden wurden bei der Nutzung des FAS beobachtet, wobei die VP explizit dazu aufgefordert waren, ihre Gedanken und Eindrücke zum System und zur Systemnutzung offen zu artikulieren (sog. „Thinking-Aloud“). Aussagen und Verhalten der Probanden wurden während der Fahrt von einem ausschließlich für diese Aufgabe zuständigen Versuchsleiter protokolliert. Gegenstand der nachfolgenden Analysen sind die Aussagen der VP während der 30-minütigen Eingewöhnungsfahrt mit dem System. Insgesamt wurden 599 Aussagen gesammelt und in die Auswertung einbezogen. Die Aufzeichnung der Beobachtung und Vorbereitung der Daten erfolgte hierbei zunächst nicht im Hinblick auf die Erstellung eines Affinitätsdiagramms.

3.2 Prozedur

Vor dem Beginn der eigentlichen Kategorisierung mit Hilfe des Affinitätsdiagramms wurde das vorliegende Datenmaterial digitalisiert und bereinigt: Aussagen wurden hinsichtlich Rechtschreibung und Grammatik korrigiert, Aussagen ohne erkennbaren Themenbezug wurden aus dem Datensatz entfernt und komplexe Mehrfahraussagen wurden in einzelne Teilaussagen zerlegt. Das gesamte verbleibende Material von 599 Aussagen wurde sodann computergestützt auf Karteikarten aufgebracht (siehe Abbildung 2). Zu Beginn der eigentlichen Kategorisierung wurden alle Aussagen nebeneinander auf einer Wand angebracht, für die Kategorisierung selbst stand eine Fläche von ca. 10 m² zur Verfügung, was die simultane Inspektion aller Karten erlaubte. Sodann wurden inhaltlich ähnliche und zusammenhängende Aussagen identifiziert und räumlich gruppiert (Spool, 2004). Hierbei ist es wichtig, jede einzelne Aussage (gelb) auf die zu Grunde liegende Intention des Nutzers zu analysieren. Der nächste Schritt bestand darin, für die verschiedenen Gruppen von Aussagen Überschriften (blau, 1. Ebene) zu finden, die alle darunterliegenden Aussagen ausreichend

beschrieben. Diese sollten aus der Sicht des Nutzers formuliert werden um die „Stimme des Nutzers“ zu verkörpern (Mazur, 2003), also z. B.: „Das Beschleunigungsverhalten des Systems ist mir zu stark“. Alle darunterliegenden Aussagen müssen sich unter dieser Aussage subsumieren lassen, ohne dass Information verloren geht. Im Anschluss konnten diese Überschriften durch eine weitere Übergruppierung (pink, 2. Ebene) zusammengefasst werden. Abschließend wurde noch eine weitere abstraktere, übergeordnete Ebene (grün, 3. Ebene) gebildet, um auf eine überschaubare Oberkategorisierung zu gelangen – als Orientierungspunkt sollte hierbei der Richtwert von maximal 7 +/- 2 Kategorien angepeilt werden, welcher aus der Forschung zur Kapazität menschlicher Reizaufnahme bekannt ist (cf. Miller, 1956). Zur Weiterverarbeitung wurden die Resultate dem digitalen Aussagenkatalog wieder zukodiert.

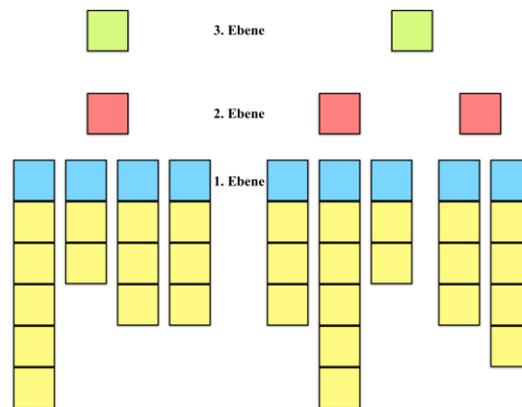


Abbildung 2. Modell eines Affinitätsdiagramms. Gelbe Post-its = Aussagen der Nutzer; Blaue Überschrift = 1. Ebene; Blaue Überschrift = 2. Ebene; Grüne Überschrift = 3. Ebene

4 Ergebnisse

Während des Kategorisierungsvorgangs wurden 9 Aussagen aufgrund der schlechten Datenqualität als nicht kategorisierbar eingestuft, so dass 590 Aussagen im Affinitätsdiagramm verblieben. Das Affinitätsdiagramm des vorliegenden Datenmaterials ergibt 8 übergeordnete Teilgebiete (grün, 3. Ebene). Ebene 2 bestand aus 13 Kategorien. Ebene 1 umfasste 65 subsumierende Aussagen. Die Verteilung der Aussagen über die verschiedenen Teilgebiete ist im Folgenden aufgezeigt (siehe Abbildung 3).

Da jeder Proband sich während der Fahrt mit dem FAS auch grundsätzlich mehrfach zur gleichen Kategorie äußern konnte, wurden weitergehende Analysen zur Struktur des Affinitätsdiagramms durchgeführt: So wurden für jede Kategorie die Anzahl der Aussagen und die Anzahl der Probanden, welche die Aussagen zu einer Kategorie abgegeben haben, betrachtet. Denkbar wäre z.B., dass nur ein einzelner Proband sehr häufig Aussagen zu einer Kategorie geäußert hat, so dass diese Kategorie letztlich nicht repräsentativ für die Stichprobe

bzw. Erhebung ist. In Abbildung 4 ist die Verteilung der beiden beschreibenden Variablen zu sehen. Beide Verteilungen ähneln sich sehr, was dafür spricht, dass es hier keine derartigen Ausreißer gab, welche die subjektive Meinung des Einzelnen durch viele Aussagen unverhältnismäßig oft in die Kategorie hat einfließen lassen.

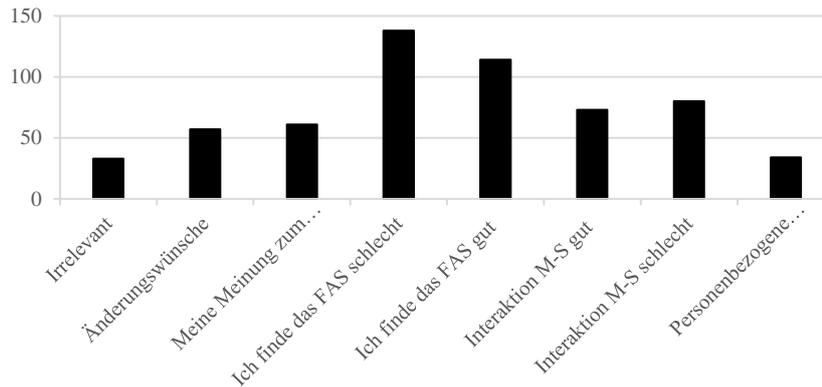


Abbildung 3. Aufteilung der 590 Aussagen auf die acht Kategorien.

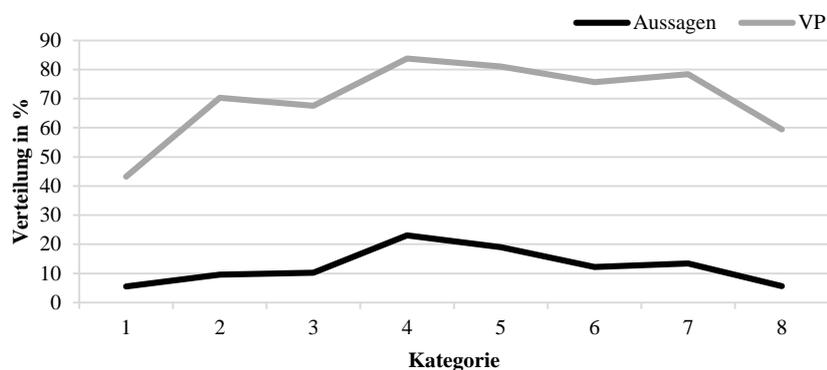


Abbildung 4. Verteilung der Anzahl der Aussagen und der Anzahl der VP, die Aussagen zu einer Kategorie getroffen haben.

5 Diskussion

Die Ergebnisse der Auswertung mit Hilfe eines Affinitätsdiagramms zeigen detaillierte Bedürfnisse des Nutzers sowie eine themenbezogene Kategorisierung der Daten. Durch das angewandte Schema lassen sich die Daten je nach Nutzen aufbereiten und verwerten. Im Vergleich zur Auswertung mit den Schemata nach Banks und Stanton (2015; 2016) mit

vorgegebenen, inhaltlich abgeleiteten Kategorien ergibt sich eine deutlich detailliertere Datenklassifizierung. Nach Banks und Stanton (2016) ergab sich z. B. nach Anwendung auf das vorliegende Datenmaterial eine Kategorie *Automation in Control* mit insgesamt 187 Aussagen, was gut 20 % aller Aussagen sind – eine detailliertere Darstellung findet aber nicht statt. Dies macht es unmöglich eine Aussage zu dieser Kategorie zu treffen, da zu viele verschiedene Aussagen in dieser Kategorie zusammengefasst wurden. Das Modell eines Affinitätsdiagramms ermöglicht hingegen die ständige nachträgliche Gruppierung in eine übergeordnete Kategorie, so dass auf der letzten, übergeordneten Ebene eine überschaubare und übersichtliche Zahl an Kategorien vorliegt. Je näher sich die Kategorien an den Daten befinden, desto mehr Aussagekraft liegt hingegen in den Überschriften, wobei die 1. Ebene gewissermaßen direkt die „Stimme des Nutzers“ widerspiegelt. Hier werden alle darunterliegenden Aussagen zu einer Kategorie zusammengefasst ohne dass es zu einem Informationsverlust kommt. Diese Ebene der Kategorisierung fehlte in den beiden vorgestellten top-down Kategorisierungsmethoden vollkommen. Weitere Übergruppierungen sind je nach Wunsch und Nutzen so zu gestalten, um die notwendige Aussagekraft für eine spezifische Fragestellung zu erlangen. Allerdings hat auch das Modell eines Affinitätsdiagramms seine Grenzen. Bei einer zu hohen Datendichte (über 1000 Aussagen) wird das genannte Modell ineffizient. Dort lohnt es sich auf bereits bestehende Kodierschemata aus Affinitätsdiagrammen zurückzugreifen, die in diesem Bereich schon an anderem Datenmaterial generiert wurden. Alternativ könnten die Kategorien vorläufig aus einer zufälligen Teilstichprobe der Daten entwickelt werden.

Durch die Anwendung eines Affinitätsdiagramms auf die vorliegenden Daten wurde ein passgenaues Kodierschema erzeugt, das einen breiten und zugleich verlustfreien Überblick über die erhobenen Aussagen verschafft. Mit Hilfe dieser Ergebnisse fällt es leicht, Aussagen über Akzeptanz, Erleben und Umgang des Probanden mit dem FAS zu treffen. Es lassen sich positive sowie negative Tendenzen erkennen und sehr unterschiedliche Meinungen zum System abbilden. Die Häufung an gleichen Aussagen von unterschiedlichen Probanden ist ein Zeichen für die Gewichtung der Aussage. Je mehr gleiche Aussagen von unterschiedlichen Probanden zu einem Thema genannt werden, desto wichtiger ist es sich mit genannter Aussage zu beschäftigen, falls sie eine negative Tendenz aufweist. Bei einer positiven Tendenz lassen sich die Stärken des Systems gut betonen. Die Klassifizierung stellt somit einen guten Ausgangspunkt für Systemüberarbeitungen dar. Eine Hürde des Modells ist die Erstellung des Affinitätsdiagramms selbst. Jede Aussage des Probanden sollte auf die dahinterliegende Intention heruntergebrochen oder aufgerundet werden. Daher ist die Qualität der erhobenen Daten sehr wichtig und kann den Prozess der Erstellung des Affinitätsdiagramms sehr erleichtern oder erschweren. Bei der Arbeit mit offenen Interviews oder Beobachtungen sollten diese schon im Hinblick auf die Erstellung eines Affinitätsdiagramms erhoben werden; Aussagen sowie Handlungen jedes Probanden unterschieden, unterteilt und auf Post-its zur weiteren Verwendung geschrieben werden. Die Unterscheidung der Variablen Anzahl der Aussagen und Anzahl der VP pro Kategorie ist wichtig, um Verzerrungen der Quantifizierung durch einzelne Probanden vorzubeugen.

Literaturverzeichnis

- Banks, V. A., & Stanton, N. A. (2015). Discovering Driver-vehicle Coordination Problems in Future Automated Control Systems: Evidence from Verbal Commentaries. *Procedia Manufacturing*, 3, 2497-2504.
- Banks, V. A., & Stanton, N. A. (2016). Keep the driver in control: Automating automobiles of the future. *Applied ergonomics*, 53, 389-395.
- Baxter, K., Courage, C., & Caine, K. (2015). *Understanding Your Users: A Practical Guide to User Research Methods*. Morgan Kaufmann.
- Beyer, H., & Holtzblatt, K. (1997). *Contextual Design: Defining Customer-Centered Systems*. Elsevier.
- Gärtner, M., Meschtscherjakov, A., Maurer, B., Wilfinger, D., & Tscheligi, M. (2014, September). Dad, stop crashing my car!: Making use of probing to inspire the design of future in-car interfaces. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 1-8). ACM.
- Holtzblatt, K., Wendell, J. B., & Wood, S. (2004). *Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design (Interactive Technologies)*. San Francisco: Morgan Kaufmann.
- Kawakita, J. (1967). *Method for idea generation. For the Development of creativity* (in Japanese). Chuo-Koron-sha.
- Kuniavsky, M. (2003). *Observing the User Experience: A Practitioner's Guide to User Research*. Morgan Kaufmann.
- Mazur, G. (2003, January). Voice of the customer (define): QFD to define value. In *ASQ World Conference on Quality and Improvement Proceedings* (Vol. 57, p. 151). American Society for Quality.
- Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review*, 63(2):81-97.
- Pyzdek, T. (2001). *The Six Sigma Handbook. A Compete Guide for Greenbelts, Black Belts and Manager at all Levels*, McGraw-Hill, New York.
- Rodríguez, M. D., Ibarra, J. E., Roa, J. R., Curlango, C. M., Bedoya, L. F., & Montes, H. D. (2014). Ambient Gamification of Automobile Driving to Encourage Safety Behaviors. In *Ubiquitous Computing and Ambient Intelligence. Personalisation and User Adapted Services* (pp. 37-43). Springer International Publishing.
- Shafer, S. M., Smith, H. J., & Linder, J. C. (2005). The power of business models. *Business horizons*, 48(3), 199-207.
- Spool, J. M. (2004). The KJ-technique: A group process for establishing priorities. *User interface engineering*.
- Wu, Y., Tan, H., & Zhao, J. (2013). Exploring aesthetics factors of gesture interaction in vehicles: An empirical study of in-car music player. *Proceedings of 5th International Automotive User Interface and Interactive Vehicular Applications*.