



Hilmar Woyczehowski

# Mit allen Sinnen

Multimodalität macht das World Wide Web noch lebendiger



Analysten beurteilen das Potenzial von multimodalen Konzepten skeptisch. Dabei bieten sie eine Reihe von Anwendungsszenarien für die Interaktion im Web. Service Providern ermöglichen sie den Zugang zum interaktiven Markt.

Das Web 3.0 bildet noch nicht die letzte Stufe der Web-Interaktivität. Gegenwärtig ist das Web primär auf Bild und Ton und damit auf Sehen und Hören oder beides zusammen ausgerichtet. Multimodale Konzepte, die über das Hören und Sehen hinausgehen, entwickeln die semantische Grundlage und die Bedeutung des Kontextes zwischen den Bits im Web 3.0 weiter. Konkret meint multimodale Kommunikation, auch menschliche Sinne wie Schmecken, Riechen, Bewegungsbewusstsein und sogar die Vorstellungskraft für neue Kommunikations- und IT-Dienste zu nutzen. Ausgelöst wird diese Entwicklung durch den enormen Fortschritt der Maschine-Mensch-Schnittstelle im Rahmen der Neurobotics-Forschung sowie alternativen Technologien, die diese Informationen transportieren.

Im Zusammenhang mit dem semantischen Web und Web 3.0 erscheinen insbesondere die mit dem Bewusstsein und der Vorstellungskraft verknüpften Sinne interessant, um ein „Smart Web“ zu entwickeln und die Vorzüge der Multimodalität tatsächlich nutzen zu können. Gängige Markttrends im Bereich multimodaler Interaktion sind die Etablierung intuitiver Common User Interfaces (CUI), die die gesamte Bandbreite der menschlichen Sinne nutzen, sowie die allmähliche Produktdifferenzierung, zum Beispiel Dolby 7.1 oder Nintendo Wii, und schließlich die Verbesserung der Sicherheit durch zusätzliche Informationskanäle.

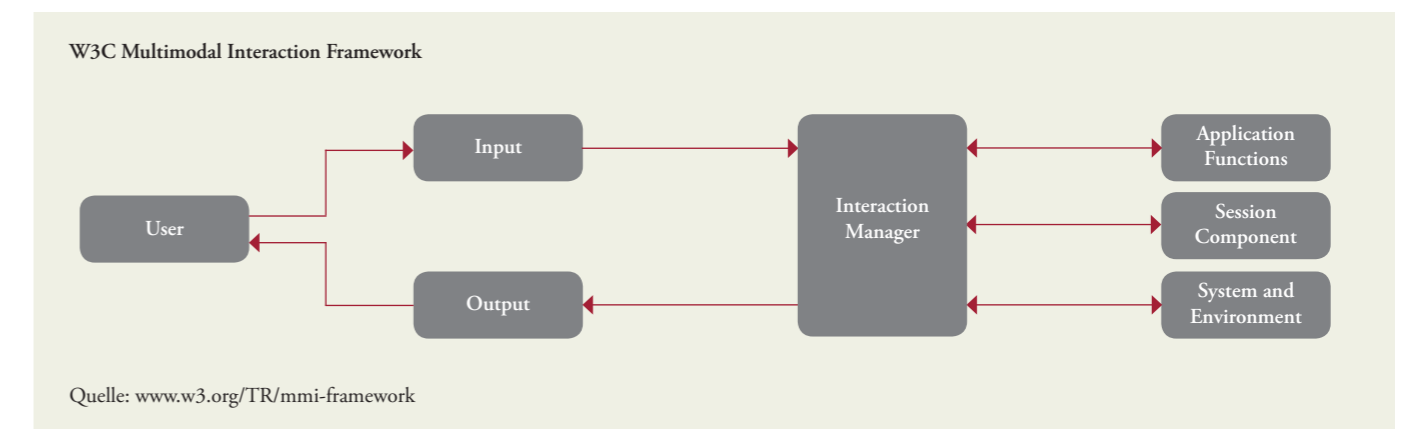
Die neueste Generation der Spielekonsolen verwendet bereits multimodale Konzepte mit neuen Tools, die eine bessere Kontrolle des Programms gewährleisten – seien es Datenhandschuhe, die die Bewegung der Hand erfassen und ein Berührungsfühl übermitteln, oder Bedienungspulte, die das Bewegungsbewusstsein weiterleiten.

Die Kontextsensitivität und die semantische Komponente innerhalb des Web 3.0 tragen in erheblichem Maße zur Verbesserung der Steigerung der Vorstellungskraft bei und führen zu einer zielorientierten und sinnvollen Interaktion mit dem Web, ohne den Output für relevante Teile zu filtern.

## Standards demonstrieren Ernsthaftigkeit

Die Standardisierungsbemühungen für multimodale Kommunikation führt das W3C Multimodal Interaction Framework und dessen nachfolgende Spezifikationen an (<http://www.w3.org/TR/mmi-framework/>). Im Mittelpunkt stehen zwei konzeptionelle Eckpfeiler für multimodale kontextsensitive Dienste für Endverbraucher: Ein Multimodal Interaction Framework gemäß der Definition des W3C, der den Service-Zugang über die unterschiedlichen Modalitäten für unterschiedliche Echtwelt-Situationen und -Ziele zur Verbesserung der Interaktion und Kommunikation mit oder zwischen den Endgeräten ermöglicht, und das situative „egozentrische Interaktionsmodell“ zur Einrahmung der Interaktion mit Objekten in der Nähe eines individuellen Nutzers im physischen und virtuellen Raum und ob diese manipulierbar sind oder nicht. Das Modell erstreckt sich ebenfalls auf andere Echtwelt- und/oder Rechneinheiten als nur das aktuell genutzte Gerät selbst.

Die Basiskomponenten des Interaction Framework sind in der folgenden Abbildung dargestellt, in der Input und Output mehrere Arten der Kommunikation und Interaktion mit dem menschlichen Nutzer widerspiegeln. Der Kern des Multimodal Interaction Framework ist der Interaktionsmanager, der die Daten und die Ausführungsströme über Input- und Output-Modalitätskomponenten koordiniert und lenkt sowie den



Status der Interaktion und den Kontext der jeweiligen Anwendung über externe Komponenten aufrechterhält.

Die Input-Komponente befasst sich mit der Erkennung, Interpretation und Integration des eingehenden Kommunikationsstroms. Die Output-Komponente auf der anderen Seite setzt sich zusammen aus Generierung, Styling und Rendering sowie Arbeitsschritten während der Rückgabe an den Endverbraucher. Kommunikationsarten können zum Beispiel Sprach-, Audio-, Schreib-, Grafik-, Video- und externer Geräte-Input sein.

Beim egozentrischen Interaktionsmodell wird die Klassifizierung der Subspaces und Objekte aus der speziell wahrnehmenden und kognitiven Erfahrung des menschlichen Akteurs im Hinblick auf die aktuelle Umgebung abgeleitet. Das Computersystem in einem Multimodal Interaction Framework sollte sich an diese wahrnehmenden und kognitiven Modelle anpassen, die menschliche Akteure als Referenzzentrum für die Maschinendarstellung betrachten. Die Modellbildung der Interaktion sollte – insbesondere in Verbindung mit semantischen Funktionen innerhalb einer Web 3.0-Umgebung – auf diesem Referenzmodell basieren und diesem nachgebildet sein.

Einige der Unterkategorien des W3C Framework wurden bislang detailliert dargestellt. Zwei oder mehr Output-Komponenten bilden eine Multimedia-Architektur, während zwei oder mehr Input-Komponenten eine multimodale Architektur darstellen. Funktionen können über mehrere Komponenten oder Module der Architektur hinweg geteilt oder zusammen in einem einzigen Element innerhalb der Architektur angeordnet werden.

Da der gegenwärtige Entwicklungsstand der multimodalen Interaktion die Rahmenbedingungen skizziert und nicht eine detaillierte Systemarchitektur darstellt, besteht immer noch Spielraum für individuelle Interpretationen und Ableitungen durch Hersteller oder Forschungsinstitute.

#### **Vielfalt der Anwendungen erhöht Relevanz auf dem Markt**

In den vergangenen Jahren gab es bereits einige internationale Forschungsprojekte rund um das Thema multimodale Kommunikation. Hervorzuheben sind zum Beispiel SmartKom des Deutschen Forschungszentrums für Künstliche Intelligenz ([http://smartkom.dfki.de/start\\_de.html](http://smartkom.dfki.de/start_de.html)) oder Mobilife (<http://www.ist-mobilife.org/>), einem Konsortium bestehend aus 22 Partnern – vom Betreiber bis zum Hersteller und wissenschaft-

lichen Partner – in neun Ländern als Teil der Wireless World Initiative (WWI).

Insbesondere in Verbindung mit der Semantik lassen sich mehrere Anwendungsfälle um die multimodale Interaktion herum konstruieren. Die sequenzielle Fähigkeit der Geräte beschreibt das automatische Hin- und Herschalten des Geräts zwischen den unterschiedlichen Arten, je nachdem, was für die Erfüllung der Anforderung des Nutzers am geeignetsten ist. Ein Mobilgerät kann zum Beispiel zwischen sprachlicher und visueller Darstellung in Bezug auf die Umstände und Präferenzen des Nutzers hin- und herschalten. Oder der Nutzer füllt ein Formular mittels eines Interviews mit dem Gerät aus, in dem nur nach den relevanten und geforderten Informationen gefragt wird. Zum Schluss stellt das Gerät das ausgefüllte Formular bildlich zur Nachbearbeitung dar (Formularausfüllung). Ein anderes Szenario beschreibt virtuelle Konferenzräume, das heißt der Nutzer erhält zusätzliche Informationen über Teilnehmer, Agenda, Diskussionspunkte, Sitzungsprotokolle und Kontaktinformationen in unterschiedlichen Output-Arten, während er an einem Meeting oder einer Telefonkonferenz teilnimmt. Außerdem verfügt der Konferenzvorsitzende über administrative Rechte, um Teilnehmer zuzulassen oder auszuschließen oder Informationen herauszugeben. Schließlich sind Mehrfach-Geräteanwendungen denkbar, das heißt Anwendungen laufen über Mehrfach-Geräte unter Verwendung unterschiedlicher Betriebsarten. Für dieses Szenario ist die Synchronisierung der Sitzungszustände bei allen Geräten von zentraler Bedeutung, was eine zentrale Netz-Entity beim Service-Provider, also dem Interaktions-Manager, erfordert.

Mit der Verbindung von multimodaler Interaktion innerhalb eines Gerätes oder über viele Geräte hinweg wird die Vision von der Kommunikation jederzeit, überall und auf jedem Gerät realistischer. Multimodale Kommunikation und Dienste können dann in der Mobilkommunikation in Verbindung mit Ort, Anwesenheit, Signalstärke, umgebendem Geräuschpegel, Gerätefähigkeiten und Monitore oder Batterieanzeige entstehen oder im Bereich Fahrzeuge mit multimodalen On-Board-Fähigkeiten, um Dienste an die Fahrersituation anzupassen, zum Beispiel Ort, Route, Interpretation der aktuellen Position, Karteninteraktion, Maschinensensoren und Benachrichtigungen, Geräuschpegel oder externe Kommunikation. Auch Desktop-Anwendungen, die statische oder dynamische Nutzerpräferenzen (automatische Anpassung) in Bezug auf Input- und Output-Komponenten verwenden, sind relevant.

Noch haben die Forschungs- und Entwicklungsbemühungen der vergangenen Jahre nicht zu einer beständigen und breiten Übernahme dieses Ansatzes in Richtung Semantic oder Smart Webs geführt. Doch mit der starken Wertaussage für diese Entwicklung sollte die Evolution des Web 3.0 nicht auf dem Radarschirm ignoriert werden. Mit dem multimodalen Rahmen und dem zentralen Interaktionsmanager ergibt sich für Service Provider die interessante Möglichkeit, auf den interaktiven Markt zu gelangen und neue vielversprechende Mehrwertdienste für die Zukunft anzubieten, die Mehrfach-Input- und Output-Modi vereinen.

Es wird kein allgemeingültiges Szenario geben, das von jedem übernommen wird, oder einen einzigen Weg, dem jeder Betreiber oder Service Provider folgt. Das World Wide Web wird mit der Übernahme semantischer und multimodaler Konzepte, bei denen jeder Player seine Position bestimmen muss, künftig noch farbenfroher und vielfältiger.

Hilmar Woyczehowski arbeitet als Consultant in der Competence Practice „Communication Technology“ bei Detecon in der Gruppe Technology Strategy. Hier beschäftigt er sich mit Themen an der Schnittstelle von Technologie und Geschäftsstrategie im Hinblick auf die zunehmende Konvergenz von Telco und IT. Er berät nationale und internationale Unternehmen in den Bereichen Technologiebewertung, Migrationsstrategien, Konzeption von Geschäftsmodellen für innovative Dienste und Evolution zu Next-Generation Networks.

Hilmar.Woyczehowski@detecon.com

