

Eine multimodale Eingabeagentur

Britta Lenzmann

Technische Fakultät, Universität Bielefeld, 33501 Bielefeld
britta@techfak.uni-bielefeld.de

Ipke Wachsmuth

Technische Fakultät, Universität Bielefeld, 33501 Bielefeld
ipke@techfak.uni-bielefeld.de

Zur Verbesserung der Intuitivität in der Mensch-Maschine-Kommunikation werden in den letzten Jahren Systeme entwickelt, die multimodale Eingabeinformationen aufnehmen und verarbeiten können. Dabei sind sprachliche und gestische Eingaben von zentraler Bedeutung, wobei eine der Hauptaufgaben – neben der Sensordatenanalyse – in der Integration der analysierten Sensorinformation besteht. Eine Integration durch reine Ausnutzung von temporalen Nachbarschaftsrelationen und sprachlich-deiktischen Verweisen erlaubt die Handhabung unvollständiger sprachlicher Anweisungen nur schwerlich. Daher verfolgen wir einen Ansatz, bei dem die Integration durch Verarbeitung in Zeitzyklen erfolgt.

Hierfür haben wir eine *multimodale Eingabearchitektur* (MEA) entwickelt; sie umfaßt (1) mehrere Eingabeagenten zur sensorischen Aufnahme, Analyse und Aufbereitung der sprachlichen und gestischen Eingaben und (2) einen Koordinatoragenten zur Integration der analysierten Sensorinformationen. Während die Eingabeagenten rhythmisch nach Eingabeereignissen “horchen”, arbeitet der Koordinator in Zeitzyklen fester, längerer Dauer. Diese Vorgehensweise ist durch menschliche temporale Kontrollmechanismen motiviert, die unterschiedliche Sensorinformationen dann als zusammengehörig interpretieren, wenn sie in einem gleichen Zeitabschnitt fester Dauer wahrgenommen werden [1]. Dementsprechend sammelt der Koordinator während eines Zeitzyklus von Sprach- und Gestenagenten kommende Informationen auf und versucht sie zu einer Gesamteingabe zu integrieren. Dabei werden in zyklischer Reihenfolge folgende Funktionen durchgeführt: **sense** zur Aufnahme analysierter Sensorinformationen, **puffer** zu deren Speicherung in einer Eingabedatenstruktur (EDS), **compute** zu deren Integration zu einer multimodalen Gesamteingabe und **act** zu deren Weitergabe an zuständige Agenten des restlichen Agentensystems. Während **sense** und **puffer** solange fortgesetzt werden, bis der aktuelle Zeitzyklus beendet ist, werden **compute** und **act** am Ende jedes Zeitzyklus ausgeführt.

Die Integration innerhalb von **compute** erfolgt durch (1) die Anzahlbestimmung von erfolgten Gesten und potentiellen Gestenplätzen, (2) der Bewertung der Gestenplätze und (3) der Zuordnung der Gesten zu den Gestenplätzen. Als potentielle Gestenplätze fungieren Spezifikationen von in den Sprachanweisungen vorkommenden Referenzobjekten sowie ortsbezeichnende Spezifikationen. Die Bewertung von Gestenplätzen wird nach der Heuristik “Die Bewertung ist umso höher, je ambiger Referenzobjekte oder Lokalisationen in der Sprachanweisung beschrieben werden.” vorgenommen. Die Zuordnung der Gesten zu den Gestenplätzen findet in Abhängigkeit von den Bewertungsergebnissen und dem Verhältnis zwischen den beiden Anzahlen statt. In der jetzigen ersten Implementation dieses Verfahrens liegt die Zyklusdauer bei 5 Sekunden. In Zukunft soll einerseits die Integration innerhalb eines Zeitzyklus verbessert werden, andererseits die Frage einer zyklusübergreifenden Integration untersucht werden.

Literatur

- [1] E. Pöppel und D. Schwender. Temporal Mechanisms of Consciousness. *International Anesthesiology Clinics* 31(4), 1993, 27-38.