Deutschsprachige Kurzfassung der Dissertation

Schritthaltende Sprachdialogverarbeitung: Architektur und signalnahe Komponenten

Timo Baumann

Diese deutschsprachige Kurzfassung meiner Dissertation "Incremental Spoken Dialogue Processing: Architecture and Lower-level Components" (Baumann 2013) stellt Ziel, Arbeitsgegenstand und die erreichten Ergebnisse dar und gibt eine kurze Übersicht über die einzelnen Kapitel und die wesentlichen wissenschaftlichen Beiträge.

In meinem Promotionsprojekt habe ich mich mit einer der technischen Beschränkungen von bisherigen Sprachdialogsystemen beschäftigt, die sich aus dem für diese üblichen Verarbeitungsschema ergeben.

Sprachdialogsysteme (SDS), als die bedeutendste Anwendung interaktiver Sprachdialogverarbeitung, sind Mensch-Maschine-Schnittstellen, bei denen gesprochene Sprache die primäre Form der Interaktion darstellt. Sprachdialogsysteme finden sich als *intelligente Assistenten* wie Apple's Siri oder Google Now neben den früher vorallem verbreiteten serverbasierten Systemen in jüngerer Vergangenheit auch auf Smartphones. Die Nutzung von gesprochener Sprache zur Interaktion mit Smartphones ist überraschend, weil diese durch vollflächige Touchscreens eigentlich ideale Bedingungen für die visuo-taktile Interaktion bieten. Aus diesem Umstand kann auf eine zukünftig noch steigende Relevanz von Sprachdialog für die Interaktion mit intelligenten Assistenten geschlossen werden, insbesondere wegen der weiter anhaltenden Miniaturisierung von Computern.

Die Attraktivität gesprochensprachlicher Interaktion beruht zum einen auf der hohen Intuitivität gesprochener Sprache, zum anderen darauf, dass gesprochene Sprache eine in der Mensch-Maschine-Interaktion wenig ausgelastete Modalität nutzt. Die Natürlichkeit und intuitive Benutzbarkeit von heutigen SDSen ist zwar noch sehr beschränkt. Dennoch reichen sie trotz ihrer Einschränkungen für einige Anwendungsfälle offenbar bereits aus, wo sie nützlich sind um relativ simple Aufgaben zu lösen, wie zum Beispiel Terminvereinbarungen oder Ticketbuchungen. Die Interaktion mit einem SDS ist aber weder so unkompliziert noch so problemlos, wie

mit einem menschlichen Gegenüber. Obwohl gerade auf der funktionalen Ebene enorme Fortschritte (zum Beispiel der Spracherkennung) erzielt wurden, bleibt die Interaktions*qualität* gering. *Konversational kompetentere* Dialogsysteme könnten weitere Anwendungsbereiche erschließen, bei denen die Interaktionsqualität selber eine stärkere Rolle spielt, wie zum Beispiel beim Entertainment oder in Spielen.

In einer Usability-Studie haben Ward u. a. (2005) sieben Hauptgründe für die häufig mangelnde Gebrauchstauglichkeit von SDSen gefunden, von denen drei (Timeouts, Responsivität und Feedback) direkt auf die *Art und Weise* der Dialogführung zurückgeführt werden können (im Unterschied zu inhaltlichen Schwierigkeiten, z.B. aufgrund von Verständnisproblemen). Den drei genannten Faktoren ist gemein, dass sie auf dem übermäßig vereinfachenden Verarbeitungsschema der für SDSe üblichen *Ping-Pong-Interaktion* beruhen: Bisherige Dialogsysteme erwarten einen vollständigen und abgeschlossenen Redebeitrag, auf den das System (nach einer gewissen Verarbeitungszeit) mit einem gleichfalls vollständigen, oft ununterbrechbaren Redebeitrag antwortet. Hinreichend lange Timeouts sind notwendig um den Nutzer nicht vor Ende seines Redebeitrags zu unterbrechen, wenn sein Redefluss während der Äußerung kurz stockt. Die Timeouts in Verbindung mit Verarbeitungszeiten führen zu einer niedrigen Responsivität, die zum Beispiel für teilüberlappende Feedback-Äußerungen nicht ausreicht. Auch selbstkann ein konventionelles System auf Feedback nicht während der eigenen Äußerung eingehen, sondern dieses nur entweder ignorieren oder als Unterbrechung interpretieren und die eigene Äußerung abbrechen. Die Redebeiträge von System und Nutzer wechseln sich mit kurzen Pausen ab, was nicht den tatsächlichen Gegebenheiten natürlichsprachlicher Interaktion entspricht, die von einem beiderseitigen Geben und Nehmen lebt, und bei der auch der jeweilige Zuhörer hilft, den Redebeitrag des jeweiligen Sprechers durch Mimik, kurze Einwürfe, und dergleichen mitzugestalten. Heutige Systeme sind also für eine hohe Interaktionsqualität nicht hinreichend interaktiv; die Rückkopplungsschleife zum Anwender ist zu lang.

Das **Ziel der Arbeit** ist, ein alternatives Verarbeitungsschema für SDSe auf seine Verbesserung der Interaktionsqualität zu untersuchen, welches die Hindernisse der Ping-Pong-Interaktion überwindet. Bei dem betrachteten Verarbeitungsschema handelt es sich um schritthaltende Verarbeitung.

Schritthaltende, bzw. *inkrementelle Verarbeitung* beschreibt ein Konzept, bei der die Verarbeitung bereits während der Eingabephase abläuft (Levelt 1989) und Zwischenergebnisse bereits vor Abschluss der Eingabe erzeugt werden (Guhe 2007).[1] Im natürlichsprachlichen Dialog verläuft sowohl die menschliche Sprachproduktion (Levelt 1989) als auch -perzeption (Tanenhaus u. a. 1995) inkrementell, das heißt Menschen hören und verstehen kontinuierlich und sind in der Lage, ihre eigenen

[1] Vgl. auch die ausführliche Diskussion in Abschnitt 3.1.1 der Dissertation.

Redebeiträge während des Sprechens mit nur kurzer Verzögerung veränderten Bedingungen anzupassen. Im Prinzip agiert auch jedes konventionelle SDS inkrementell, da sich Produktion und Perzeption ganzer Redebeiträge zwischen System und Benutzer abwechseln; ihre *Granularität*, also die minimalen Verarbeitungseinheiten liegt bei ganzen Redebeiträgen. Hier wird eine wesentlich feingliedrigere Verarbeitung angestrebt, bei der, je nach linguistischem Abstraktionsgrad, die Granularität teilweise nur wenige Millisekunden gesprochener Sprache umfasst, und damit eine deutlich flexiblere Verarbeitung ermöglicht.[2]

Unter anderem haben Aist u. a. (2007a) Vorteile von inkrementellem Verstehen gegenüber nicht-inkrementellem Verstehen in der Mensch-Maschine-Interaktion gezeigt und Aist u. a. (2007b) konnten Effizienzgewinne für ein SDS mit dieser Technik nachweisen. Diese Arbeit behandelt feingliedrig inkrementelle Verarbeitung als Grundlage der Dialogsystemarchitektur (und nicht nur einzelner Verarbeitungsmodule) um vollständig inkrementelle Systeme zu ermöglichen.

Der **Arbeitsgegenstand** sind also die Dialogsystemarchitektur für feingliedrig inkrementelle Verarbeitung, eine Evaluationsmethodik für inkrementelle Verarbeitung, sowie die Entwicklung und Analyse inkrementeller *Grundfertigkeiten* für Sprachdialogsysteme.

Im Hinblick auf das Ziel einer höheren Interaktionsqualität und den Arbeitsgegenstand inkrementelle Dialogsystemarchitektur und -komponenten untersucht die Arbeit folgende **Kernthese**:

> **Feingliedrig inkrementelle und pro-aktive Verarbeitung ermöglicht natürlicher (und damit erfolgreicher) interagierende Sprachdialogsysteme.**

Um die Kernthese zu prüfen muss nachgewiesen werden, dass umfassende inkrementelle Verarbeitung in einem SDS sowohl möglich als auch erfolgreich ist und im Vergleich mit nicht-inkrementeller Verarbeitung zu natürlicherer Interaktion beitragen kann. Hierfür wird neben dem System und seiner Architektur eine umfassende Evaluationsmethodik benötigt um die *inkrementelle Qualität* einzelner Module zu überprüfen.

Drei Aspekte werden für die inkrementelle Qualität als maßgebened herausgearbeitet: Eine möglichst geringe Verzögerung bei der (Teil-)Hypothesenbildung, eine möglichst geringe Zeitspanne bis zu derer die Hypothesen als zuverlässig betrachtet werden können, sowie eine möglichst geringe Wankelmütigkeit, ausgedrückt durch möglichst wenige Nachbearbeitungen der Zwischenhypothesen hin zur endgültigen Hypothese. Diese Ziele stehen teilweise in Konflikt zueinander, zum Beispiel sind frühe Hypothesen häufig besonders fehleranfällig und erfordern gegebenenfalls mehr Nachbearbeitungen als stärker verzögerte Hypothesen.

[2] In der Fachöffentlichkeit wird unter inkrementeller Verarbeitung immer diese "feingliedrig" inkrementelle Verarbeitung verstanden (z.B. Schlangen und Rieser 2011).

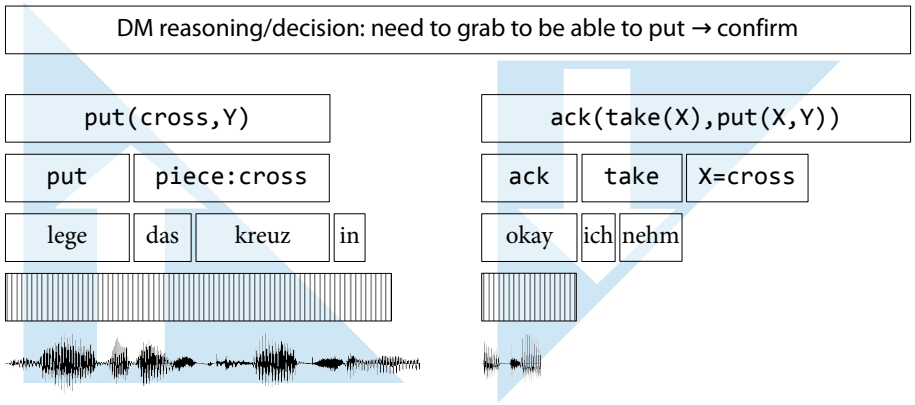

Abbildung 1: "Dreieckige" IU-Datenstrukturen in InproTK, mit bottom-up, datengetriebener Eingabeverarbeitung auf der linken und top-down, just-in-time Ausgabeerzeugung auf der rechten Seite. Das Dialogmanagement (oben) steuert die Ausgabevorbereitung auf Grundlage der Eingaben und trifft Entscheidungen, welche den Revisionsprozess der Hypothesen begrenzt.

Darüberhinaus wird dargestellt, dass Nachbearbeitungen nicht grundsätzlich auf einen bestimmten (zeitlich lokalen) Kontext beschränkt werden können, sondern Abhängigkeiten in linguistischen Daten im Prinzip unbeschränkt auftreten können.[3] Folglich muss eine Architektur für inkrementelle Verarbeitung nicht nur die Erweiterung von Hypothesen, sondern auch die unbeschränkte Abänderung und Rücknahme älterer Hypothesen unterstützen, was die entwickelte Software-Architektur, basierend auf dem IU-Modell (Schlangen und Skantze 2009), umsetzt.

Innerhalb der Architektur sind getrennte Module für die Verarbeitung auf den unterschiedlichen linguistischen Ebenen zuständig und da die Granularität mit steigendem Abstraktionsgrad tendenziell abnimmt, ergeben sich "dreieckige" Datenstrukturen, wie in Abbildung 1 idealisiert gezeigt. So erstreckt sich im Beispiel das Erkennungsergebnis noch nicht auf die nach dem Wort "`in`" folgende Spracheingabe, und auch die semantische Analyse hängt gegenüber dem Erkennungsergebnis zurück ("`in`" ist noch nicht angebunden). Auf der Ausgabeseite kann in einer abstrakten Form bereits ein (möglicherweise unterspezifizierter) Dialogakt generiert werden, ohne dass die Sprachsynthese hierfür schon umfangreich arbeiten muss (dies geschieht stattdessen erst 'just-in-time'). Das Dialogmanagement steuert die Vorbereitung von Ausgaben

[3] Das letzte Wort eines Buches kann die Bedeutung des ersten modifizieren, unabhängig von der Dicke des Buches.

schon während der Eingabe und trifft die Entscheidung für die Ausführung derselben. All dies wird über eine verlinkte objektorientierte Datenhaltung von *IUs* (in der Abbildung die einzelnen Kästchen) realisiert; Module erzeugen oder entfernen jeweils solche IU-Objekte als Reaktion auf Änderungen am IU-Netzwerk (vgl. Schlangen und Skantze 2009, 2011, sowie Kapitel 4).

Zunächst können Hypothesen beliebig korrigiert und erweitert werden. Zum Beispiel könnte als nächstes das Wort "`gelb`" erkannt werden, und angenommen die sich ergebende Referenz "`das Kreuz in gelb`" wäre in der Domäne unauflösbar. Als Folge daraus könnte das Dialogmanagement sein Nicht-Verstehen (anstatt der bisher geplanten Bestätigung) ausdrücken. Der Aufwand hierfür wäre minimiert, da bisher zum Beispiel noch kein referierender Ausdruck für `X=cross` generiert und nur ein Sprechanfang synthetisiert wurden. Für die inkrementelle Verarbeitung müssen also alle Module im Dialogsystem mit der Rücknahme bzw. Abänderung ihrer jeweiligen partiellen Eingaben und Ausgaben umgehen können, und diese Fähigkeit wird im folgenden für die signalnahen *Grundfertigkeiten* der Dialogverarbeitung implementiert.

Dafür wird innerhalb der Architektur ein Spracherkennungssystem inkrementell nutzbar gemacht (als Eingabemodul in das IU-Netzwerk) und anhand der entwickelten Evaluationsmethodik bewertet. Es zeigt sich, dass die inkrementell erzeugten vorläufigen Hypothesen sowohl in ihrer inhaltlichen Qualität als auch in ihrer Rechtzeitigkeit für SDSe nützlich sein können. Außerdem wird gezeigt, dass die prinzipielle Abwägung zwischen Hypothesensicherheit und Hypothesenverzögerung durch relativ einfache Verfahren positiv beeinflusst werden kann. Schließlich ergeben sich durch inkrementelle Spracherkennung neue Verhaltensweisen für Mensch-Computer-Interaktion die mit nicht-inkrementellen Mitteln unmöglich sind. In zwei Interaktionsstudien wird der Nutzen von direkten visuellen Rückmeldungen sowie der gemeinschaftlichen Gestaltung und Aushandlung von Äußerungen (Clark 1996) in Nutzerbewertungen gezeigt.

Zusätzlich behauptet die Kernthese, dass Entscheidungen im Dialog nicht rein reaktiv, also erst wenn der gesamte relevante Kontext zur Verfügung steht, sondern darüberhinaus *pro-aktiv* getroffen werden sollten, d.h. sobald die Chancen früher Handlungen das aus der eingeschränkten Entscheidungsgrundlage erwachsende Risiko überwiegen. Hierfür benötigen Systeme nicht nur ein Verständnis für die Vergangenheit, sondern zusätzlich ein Modell der Erwartungen der näheren Zukunft. Auch der Nutzen dieser Fähigkeit wird mit einem signalnahen Modul nachgewiesen, nämlich in einem System welches die Sprechzeit der nächsten vom Nutzer zu sprechenden Wörter (während einer bekannten Äußerung) schätzt. Diese Fähigkeit wird genutzt um zeitsynchron mit dem Nutzer dessen Äußerungen mitzusprechen und zeigt damit, dass *durchgehende inkrementelle Verarbeitung* in Echtzeit tatsächlich möglich ist: Die Verzögerungen die sich aus der Hypothesenbildung der Spracherkennung sowie der

Verarbeitungszeiten der Sprachsynthese[4] ergeben, können durch entsprechend große Vorhersagezeiträume ausgeglichen werden, sodass ein Mitsprechen möglich ist, das in seiner zeitlichen Koordination menschlicher Genauigkeit nahekommt.

Die bereits genannten *Grundfertigkeiten* Spracherkennung und Sprechverlaufsschätzung werden schließlich um eine vollständige, inkrementell nutzbare Sprachsynthesekomponente ergänzt, sodass Systemäußerungen bereits begonnen werden können, bevor sie vollständig spezifiziert sind, und bereits geplante aber noch nicht realisierte Teile der Äußerung noch abgeändert werden können. Dies erlaubt Ausgaben noch bevor der vollständige Inhalt bestimmt ist, oder das Einbeziehen von Nutzerreaktionen während einer laufenden Systemäußerung. Auch die Ausgabe steht dabei vor dem Zielkonflikt inkrementeller Verarbeitung, dass die Zeitigkeit des Verhaltens mit dem Umfang des zur Verfügung stehenden Kontextes konkurriert. In diesem Fall ist insbesondere die generierte Prosodie und damit die Sprachausgabequalität schlechter, wenn kein ausreichender Kontext zur Verfügung steht. Der resultierende Qualitätsverlust aus inkrementeller Verarbeitung wird für unterschiedlich große Kontexte quantifiziert und für relativ große Kontexte als vernachlässigbar eingestuft. Zusätzlich zeigt ein Experiment, dass die völlig neuen Interaktionsmöglichkeiten durch inkrementelle Sprachausgabe als wesentlicher Fortschritt der Interaktionsqualität bewertet werden, und die Sprachqualität subjektiv sogar als besser als die nicht-inkrementelle Variante wahrgenommen wird. Dies impliziert, dass Nutzer die Interaktionsqualität stärker gewichten als die Sprachqualität.

Insgesamt zeigt die Arbeit, dass kontinuierliche, inkrementelle Sprachverarbeitung sowohl für die Erkennung als auch die Synthese mit gutem Erfolg möglich ist. Insbesondere überwiegen die Chancen inkrementeller Verarbeitung ihre inhärenten Nachteile. Bereits im gezeigten Umfang zeigt sich, dass inkrementelle Sprachdialogverarbeitung deutliche Vorteile für die Mensch-Maschine-Interaktion bietet.

Die Arbeit fokussiert ganz bewusst nur auf den Grundfertigkeiten Hören, Sprechen, und Sprechverlaufsschätzung; Module übergeordneter Abstraktionsgrade sind in den Beispielsystemen teilweise weniger ausgefeilt, oder nur simuliert. Einige andere Arbeiten haben hierauf aufbauend bereits Verarbeitungskomponenten auf signalferneren Abstraktionsebenen erstellt (Buß und Schlangen 2011; DeVault, Sagae und Traum 2009; Peldszus u. a. 2012; Sagae u. a. 2009; Schlangen, Baumann und Atterer 2009). Die erstellten Komponenten sind also Teil eines vitalen Ökosystems inkrementeller Verarbeitung. Ferner fokussiert die Arbeit ausschließlich auf Sprachdialogsysteme. Die erstellten Modelle und Komponenten sind aber nicht dialog-spezifisch sondern können auch für ähnliche inkrementelle Anwendungsfälle nützlich sein, wie zum Beispiel maschinelles Dolmetschen (Amtrup 1999; Bangalore u. a. 2012).

[4] In diesem Experiment wird eine wortweise Sprachausgabe benutzt; ihre mangelnde Qualität führte im weiteren zur Entwicklung echter inkrementeller Sprachsynthese.

Übersicht über die Kapitel

Kapitel 1 leitet in das Thema ein, erarbeitet die Leitfrage der Arbeit und gibt eine kurze Übersicht über die bearbeiteten Teilgebiete. (Diese Kurzfassung entspricht im Wesentlichen dem Kapitel 1 der Dissertation.)

Kapitel 2 ordnet die Arbeit ein, indem es einen Überblick über Fragen der gesprochensprachlichen Interaktion, des Dialogs und bisherigen Dialogsystemen gibt.

Kapitel 3 vertieft die Thematik der schritthaltenden (inkrementellen) Verarbeitung auf Basis der Literatur und führt einen Formalismus für die Darstellung von partiellen, inkrementell erweiterbaren Hypothesen ein, anhand dessen Qualitätsmaße inkrementeller Verarbeitung definiert werden (Verzögerung bei der Hypothesenbildung, Verzögerung bis Hypothesen zuverlässig werden, sowie die Häufigkeit mit der Hypothesen nachbearbeitet werden müssen), die ausführlich diskutiert werden.

Kapitel 4 stellt die Architektur des im Rahmen der Arbeit entwickelten Softwaretoolkits für schritthaltende Verarbeitung InproTK vor und diskutiert Daten- und Verarbeitungsschemata.

Kapitel 5 betrachtet inkrementelle Spracherkennung. Die ‚inkrementelle Qualität' der Spracherkennung wird intensiv auf mehreren Korpora und für unterschiedliche Varianten in all ihren Aspekten untersucht. Schließlich werden Optimierungsmethoden vorgestellt, welche Qualitätsaspekte gegeneinander abwägen. Der Nutzen inkrementeller Spracherkennung wird beispielhaft in einer Spielanwendung gezeigt.

Kapitel 6 geht den Schritt von möglichst reaktiver zu *proaktiver* Verarbeitung, welche erlaubt, den Dialogverlauf aktiv zu steuern. Eine Beispielanwendung zeigt, wie durch schritthaltende Verarbeitung die Rückkopplung zwischen Nutzer und System beschleunigt und dadurch Nutzeräußerungen gemeinschaftlich gestaltet werden können. Schließlich wird ein System vorgestellt, welches Nutzeräußerungen synchron mitspricht. Dieses System zeigt, dass inkrementelle und proaktive Verarbeitung synchrone Interaktionsfähigkeiten in Echtzeit ermöglichen, indem alle Systemverzögerungen an anderer Stelle durch Prädiktion ausgeglichen werden. Gleichzeitig wirft dieses System den Wunsch nach echter inkrementeller Sprachsynthese auf, da der hier verfolgte Ansatz wortweiser Synthese nicht überzeugt.

Kapitel 7 betrachtet deshalb inkrementelle Sprachsynthese, bei der die Spezifikation der Äußerung noch während der Synthese erweitert oder abgeändert werden kann, sowohl was die kommunizierten Inhalte als auch die prosodische Realisation angeht. Der Nutzen dieser Fähigkeit wird in einer hochdynamischen Umgebung demonstriert, in der Inkrementalität Reaktionen ermöglicht die von Versuchspersonen als deutlich natürlicher im Vergleich zu einem nicht-inkrementellen System bewertet werden. Schließlich wird die Integration inkrementeller Sprachsynthese mit einem inkrementellenSprachgenerierungsmodul demonstriert, und der Einfluss auf die resultierende Prosodiequalität des Systems bewertet. Es zeigt sich, dass der Vorteil

der resultierenden Interaktionsqualität die geringere prosodische Qualität überwiegt und sogar subjektiv die prosodische Qualität ansteigt.

Kapitel 8 fasst die Ergebnisse der Arbeit zusammen: Feingliedrig schritthaltende Verarbeitung ist technisch möglich und so erfolgreich, dass dadurch für Sprachdialogsysteme vormals unerreichbare Interaktionsmodi ermöglicht werden (u. a. gemeinschaftliche Äußerungsgestaltung, synchrones Sprechen, Berücksichtigung von Änderungen während Systemäußerungen). Schritthaltende Verarbeitung sollte deshalb die Basis für zukünftige Sprachdialogsysteme bilden.

Wissenschaftliche Beiträge

- Eine Evaluationsmethodik für die Bewertung der Qualität inkrementeller Verarbeitung (Abschnitt 3.3.2), die Evaluationstoolbox INTELIDA die für inkrementelle Spracherkennungsergebnisse die Evaluationsmethodik umsetzt (Abschnitt 5.3), und die erzielten Ergebnisse ausführlich diskutiert (Abschnitt 5.4), und schließlich Optimierungsmethoden für den Zielkonflikt zwischen möglichst zeitnahen und möglichst zuverlässigen Zwischenergebnissen (Abschnitt 5.5);
- eine Software-Architektur für inkrementelle Sprachdialogsysteme, basierend auf dem IU-Modell (Schlangen und Skantze 2009) (Kapitel 4), mit integrierter inkrementeller Spracherkennung (Abschnitt 5.2.1) und -synthese (Abschnitt 7.3.1), sowie Dialogablaufkontrolle (Abschnitt 6.1) und Zeitverlaufsvorhersage (Abschnitt 6.2), die als freie, quelloffene Software veröffentlicht ist[5];
- eine inkrementelle parametrische Sprachsynthese mit Echtzeitmanipulation von Sprachparametern und -inhalten (Abschnitt 7.3.1) die hochgradig interaktives Systemverhalten ermöglicht (Abschnitte 7.4 und 7.5), sowie eine Analyse der inkrementell erreichbaren Prosodiequalität (Abschnitt 7.6);
- ein System welches in der Lage ist (bekannte) Äußerungen zeitsynchron mitzusprechen und damit beweist, dass durchgehende inkrementelle Verarbeitung in Verbindung mit Prädiktion in Echtzeit funktioniert (Abschnitt 6.2.6); und schließlich
- ein interaktives und inkrementell sprachgesteuertes System mit sofortigen visuellen Rückmeldungen an den Nutzer (Abschnitt 5.6), sowie ein vollständiges inkrementelles und kooperatives Dialogsystem (Abschnitt 6.1.2), derer beider Interaktionen jeweils als natürlicher und reaktiver bewertet wurden als die von nicht-inkrementellen Vergleichssystemen.

[5] Siehe `http://inprotk.sf.net`.

Literatur

Aist, Gregory, James Allen, Ellen Campana, Carlos Gomez Gallo, Scott Stoness, Mary Swift und Michael K. Tanenhaus (2007a). „Incremental Dialogue System Faster than and Preferred to its Nonincremental Counterpart". In: *Proceedings of the 29th Annual Conference of the Cognitive Science Society*. Nashville, USA, S. 761–766.

– (2007b). „Incremental Understanding in Human-Computer Dialogue and Experimental Evidence for Advantages over Nonincremental Methods". In: *Proceedings of DECALOG, the 11th International Workshop on the Semantics and Pragmatics of Dialogue*. Trento, Italy, S. 149–154.

Amtrup, Jan Willers (1999). *Incremental speech translation*. Berlin, Heidelberg: Springer. ISBN: 3-540-66753-9.

Bangalore, Srinivas, Vivek Kumar Rangarajan Sridhar, Prakash Kolan, Ladan Golipour und Aura Jimenez (2012). „Real-time Incremental Speech-to-Speech Translation of Dialogs". In: *Proceedings of NAACL-HLT 2012*. Montréal, Canada: Association for Computational Linguistics, S. 437–445.

Baumann, Timo (2013). „Incremental Spoken Dialogue Processing: Architecture and Lower-level Components". Diss. Universität Bielefeld, Germany. URN: urn:nbn:de:hbz:361-25819101.

Buß, Okko und David Schlangen (2011). „DIUM – An Incremental Dialogue Manager That Can Produce Self-Corrections". In: *Proceedings of SemDial 2011 (Los Angelogue)*. Los Angeles, USA.

Clark, Herbert H. (1996). *Using Language*. Cambridge University Press. ISBN: 978-0521567459.

DeVault, David, Kenji Sagae und David Traum (2009). „Can I Finish? Learning When to Respond to Incremental Interpretation Results in Interactive Dialogue". In: *Proceedings of the SIGDIAL 2009 Conference*. London, UK, S. 11–20.

Guhe, Markus (2007). *Incremental Conceptualization for Language Production*. Mahwah, USA: Lawrence Erlbaum Associates.

Levelt, William J.M. (1989). *Speaking: From Intention to Articulation*. Mit Pr.

Peldszus, Andreas, Okko Buß, Timo Baumann und David Schlangen (2012). „Joint Satisfaction of Syntactic and Pragmatic Constraints Improves Incremental Spoken Language Understanding". In: *Proceedings of EACL*. Avignon, France.

Sagae, Kenji, Gwen Christian, David DeVault und David Traum (2009). „Towards Natural Language Understanding of Partial Speech Recognition Results in Dialogue Systems". In: *Proceedings of NAACL-HLT 2009*. Association for Computational Linguistics, S. 53–56.

Schlangen, David, Timo Baumann und Michaela Atterer (2009). „Incremental Reference Resolution: The Task, Metrics for Evaluation, and a Bayesian Filtering Model that is Sensitive to Disfluencies". In: *Proceedings of SigDial 2009*. London, UK.

Schlangen, David und Hannes Rieser, Hrsg. (2011). *Dialogue & Discourse* 2.1. Special Issue on Incremental Processing in Dialogue. ISSN: 2152-9620.

Schlangen, David und Gabriel Skantze (2009). „A General, Abstract Model of Incremental Dialogue Processing". In: *Proceedings of the EACL*. Athens, Greece, S. 710–718.

– (2011). „A General, Abstract Model of Incremental Processing". In: *Dialogue and Discourse* 2.1, S. 83–111.

Tanenhaus, Michael K., M.J. Spivey-Knowlton, K.M. Eberhard und J.C. Sedivy (1995). „Integration of visual and linguistic information in spoken language comprehension". In: *Science* 268.5217, S. 1632–1634.

Ward, Nigel G., Anais G. Rivera, Karen Ward und David G. Novick (2005). „Root causes of lost time and user stress in a simple dialog system". In: *INTERSPEECH 2005*. Lisbon, Portugal: ISCA, S. 1565–1568.