

UASR - Von der Idee bis zur Open-Source-Ressource

Rüdiger Hoffmann¹, Matthias Wolff²

¹ TU Dresden, Professur Systemtheorie und Sprachtechnologie, 01062 Dresden,
E-Mail: ruediger.hoffmann@tu-dresden.de

² BTU Cottbus-Senftenberg, Lehrstuhl Kommunikationstechnik, 03046 Cottbus,
E-Mail: matthias.wolff@tu-cottbus.de

Einleitung

An der TU Dresden ist über mehr als ein Jahrzehnt lang ein hierarchisches Analyse-Synthese-System entwickelt worden, das die Bezeichnung UASR (Unified Approach for Synthesis and Analysis) erhielt. Das System diente zu Grundlagenuntersuchungen im Bereich kognitiver Systeme, aber auch als Plattform für die Entwicklung von Erkennungssystemen für Sprache sowie für nichtsprachliche Signale. In dem Beitrag wird besonders auf einige neuere Anwendungen eingegangen, die von eingebetteten Systemen bis zu Erkennungssystemen für fließende Sprache reichen. Da das System neuerdings auf einer Open-Source-Plattform frei verfügbar ist, wird abschließend diese Nutzungsmöglichkeit näher erläutert.

Idee und vorliegende Implementierung

Das Prinzip der „Analyse durch Synthese“ beschreibt eine traditionelle Herangehensweise in der Sprachtechnologie. Grundsatzüberlegungen der 1980er-Jahre [1] empfahlen bereits eine integrierte Betrachtung von Spracherkennung und -synthese, die sich historisch weitgehend unabhängig voneinander entwickelt hatten. In den 1990er-Jahren zeigte sich in Projekten wie Verbmobil [2] die wachsende Bedeutung von Sprachdatenbasen, die für Analyse und Synthese immer ähnlicher wurden, bis die aufkommende HMM-Synthese diese Tendenz dann auch praktisch bestätigte [3, 4].

Im Jahre 2000 wurde an der TU Dresden das Projekt UASR entwickelt, das das Konzept der Analyse durch Synthese unter Berücksichtigung der hierarchischen Struktur der Sprache untersuchen und praktisch umsetzen sollte [5, 6]. Das Projekt wurde durch drei aufeinanderfolgende DFG-Projekte und durch verschiedene Anwendungsprojekte finanziert. Ein wichtiger technologischer Aspekt war die einheitliche Implementierung der symbolischen Ebenen des Systems mit Hilfe von FST (Finite State Transducers); diese Ausbaustufe ist in Abbildung 1 wiedergegeben und wurde u. a. in [7] publiziert.

Das System diente nicht nur als leistungsfähige Entwicklungsbasis für Standardaufgaben der Spracherkennung, sondern kam auch in einer eingebetteten Anwendung mit FPGA zum Einsatz [8, 9]. Erwähnenswert ist auch, dass es als Grundlage für eine größere Zahl von Anwendungen im Bereich der Erkennung nichtsprachlicher Signale diente (Biosignale [10], Maschinengeräusche, zerstörungsfreie Prüfung [11], Musik [12] u. a.).

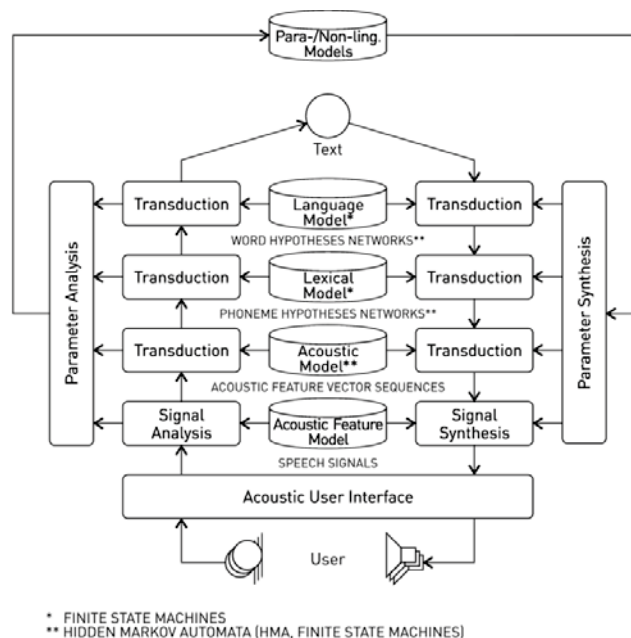


Abbildung 1: Blockschaltbild von UASR als Hierarchie aus der signalnahen, akustischen, phonetischen, lexikalischen und syntaktischen Verarbeitungsebene (von unten). Links der Analyse-, rechts der Synthesezweig.

Weiterentwicklung zum kognitiven System

Etwa in der Mitte der 2000er-Jahre entwickelte sich, ausgehend von nachrichtentechnischen Fortschritten (cognitive radar, cognitive radio [13]) eine neue Sicht auf Systeme mit intelligentem Verhalten. Haykin hat für Systeme, die wie der Mensch ein zielgerichtetes Verhalten aufweisen, die Bezeichnung *kognitive dynamische Systeme* geprägt. Sie sind durch ihre Sensoren in der Lage, ein internes Modell ihrer Umwelt zu entwickeln und auf dieser Basis gezielt auf die Umwelt einzuwirken. Analogien zu biologischen Systemen sind offensichtlich, und abgesehen von einfachen Fällen liegt eine hierarchische Struktur vor.

Aus der Verwandtschaft des UASR-Konzepts mit hierarchischen kognitiven dynamischen Systemen [14] ergibt sich das Vorhaben, UASR durch Hinzunahme einer semantischen Ebene (Dialogkomponente) auszubauen. Dazu ist die Kooperation mit Informatikern und Computerlinguisten erforderlich. Dieses interdisziplinäre Projekt hat mittlerweile Gestalt angenommen und bildete einen Schwerpunkt der Konferenz Elektronische Sprachsignalverarbeitung 2012 mit mehreren Einzelbeiträgen, auf die hier aus Platzgründen pauschal verwiesen wird [15].

UASR als Open Source

Die Quelltexte wesentlicher Teile der oben kurz beschriebenen Software-Implementierung von UASR wurden Anfang 2014 unter der GNU Lesser General Public License (LGPL) veröffentlicht [16]. Die Open-Source-Version umfasst eine Auswahl von Werkzeugen zur Sprach- und Signalverarbeitung sowie zur akustischen Mustererkennung, unter anderen:

- modellbasierte Signalanalyse und –synthese,
- HMM-Erkennung und HMM-Signalsynthese (inkl. Training, Adaption und Erzeugung HMM-kodierter Inventare für die konkatenative Sprachsynthese),
- Mustererkennung basierend auf Supportvektormaschinen¹ und Korrelationskoeffizienten,
- Erzeugung und Verarbeitung endlicher und stochastischer Grammatiken und
- allgemeine Werkzeuge (beispielsweise zur evolutionären Optimierung, Kreuzvalidierung, Datenbasisverwaltung usw.).

Die UASR-Werkzeuge sind für die Massendatenverarbeitung ausgelegt und unterstützen teilweise Parallelrechnen. UASR ist in Form von Skripten unter Verwendung einer ebenfalls 2014 unter LGPL veröffentlichten eigenen Toolbox zur Signalverarbeitung und Mustererkennung (dLabPro, [17]) realisiert. Diese umfasst neben Grundfunktionen (Algebra², Datenmanipulation, Dateizugriff, Profiling usw.):

- Signalverarbeitung (implementiert als Paare inverser Analyse- und Synthesefilter sowie Grundfrequenz- und Prosodieanalyse/-synthese),
- Statistik und Modellierung (inkl. GMM und HMM),
- Manipulation und Simulation endlicher Transduktoren (Finite State Machines) inkl. Suchalgorithmen,
- ein Software Development Kit sowie
- eine einfache objektorientierte Skriptsprache.

Die dLabPro-Toolbox ist in ANSI C (teilweise in ANSI C++) implementiert. Sie ist unter Linux, MacOS und Windows übersetzbar und getestet. Außerdem sind große Teile auf Gleitkomma-Signalprozessoren kompilierbar und lauffähig. Für einige Algorithmen existieren darüber hinaus FPGA-Layouts (XILINX Virtex, kommerzielle Lizenz).

Ebenfalls frei verfügbar ist ein Eclipse-Plugin für dLabPro und UASR, welches Editoren für dLabPro-Skripte sowie eine einfache Datenvisualisierung zur Verfügung stellt [18].

Hinweise zur Installation und Verwendung der Open-Source-Software zu UASR können in den Online-Dokumentationen unter [16-18] gefunden werden. Die Quelldatenbanken enthalten den jeweils neuesten Entwicklungsstand und die Versionsgeschichte. Die Veröffentlichung von stabilen ausführbaren Versionen (*binary releases*) von dLabPro sowie einer Testsuite ist geplant.

¹ durch eingebundene libsvm-Bibliothek [19]

² durch eingebundene LAPACK-Bibliothek [20]

Literatur

- [1] Holmes, J. N.: Speech Synthesis and Recognition. London: Van Norstrand Reinhold 1988.
- [2] Wahlster, W. (Hrsg.): Verbmobil – Foundations of Speech-to-Speech Translation. Berlin etc.: Springer 2000.
- [3] Falaschi, A.; Giustiniani, M.; Verola, M.: A hidden Markov model approach to speech synthesis. Proc. Eurospeech, Paris 1989, 187 - 190.
- [4] Tokuda, K., et al.: Speech parameter generation algorithms for HMM-based speech synthesis. Proc. ICASSP, Istanbul 2000, 1315 - 1318.
- [5] Eichner, M.; Wolff, M.; Hoffmann, R.: A unified approach for speech synthesis and speech recognition using stochastic Markov graphs. Proc. ICSLP, Beijing 2000, vol. 1, 701 - 704.
- [6] Eichner, M.; Wolff, M.; Werner, S.; Hoffmann, R.: Ein kombiniertes Spracherkennungs- / Sprachsynthesesystem auf Phonemebene. 11. Konferenz Elektronische Sprachsignalverarbeitung, Cottbus 2000, 141 - 149.
- [7] Hoffmann, R.; Eichner, M.; Wolff M.: Analysis of verbal and nonverbal acoustic signals with the Dresden UASR system. In: A. Esposito et al. (Eds.): Verbal and Nonverbal Communication Behaviours. Berlin etc.: Springer 2007 (LNAI vol. 4775), 200 - 218.
- [8] Duckhorn, F.; Strecha, G.; Wolff, M.; Hoffmann, R.: Ein Sprachdialogsystem mit begrenzten Hardwareressourcen. 20. Konferenz Elektronische Sprachsignalverarbeitung, Dresden 2009, Bd. 1, 88 - 93.
- [9] Strecha, G.; Wolff, M.; Duckhorn, F.; Wittenberg, S.; Tschöpe, C.: The HMM synthesis algorithm of an embedded unified speech recognizer and synthesizer. Proc. Interspeech, Brighton 2009, 1763 - 1766.
- [10] Wolff, M.; Kordon, U.; Hussein, H.; Eichner, M.; Tschöpe, C.; Hoffmann, R.: Auscultatory blood pressure measurement using HMMs. Proc. IEEE ICASSP, Honolulu 2007, vol. 1, 405 - 408.
- [11] Tschöpe, C.; Wolff, M.: Statistical classifiers for structural health monitoring. IEEE Sensors Journal 9 (2009) 11, 1567 - 1576.
- [12] Hübler, S.; Hoffmann, R.: Modelling drum patterns with weighted finite-state transducers. Proc. IEEE ICASSP, Vancouver 2013, 719 - 723.
- [13] Haykin, S.: Cognitive Radio: Brain-Empowered Wireless Communications. IEEE Journal on Selected Areas in Communications 23 (2005) 2, 201 - 220.
- [14] Hoffmann, R.: Denken in Systemen. In: G. Gerlach, R. Hoffmann (Hrsg.): Neue Entwicklungen in der Elektroakustik und elektromechanischen Messtechnik. Festschrift für G. Pfeifer, TUDpress Dresden 2009, 13 - 23 (Dresdner Beiträge zur Sensorik, Bd. 40).
- [15] Wolff, M. (Hrsg.): Elektronische Sprachsignalverarbeitung 2012. Tagungsband der 23. Konferenz, Cottbus, 29. - 31. 8. 2012. Dresden: TUDpress 2012 (Studentexte zur Sprachkommunikation, Bd. 64).
- [16] <https://github.com/matthias-wolff/UASR>
- [17] <https://github.com/matthias-wolff/dLabPro>
- [18] <https://www-docs.tu-cottbus.de/kommunikationstechnik/public/updatesite>
- [19] <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [20] Andersen, E., et al.: LAPACK's User Guide. Society for Industrial and Applied Mathematics, 1999.