

Rechnergestützte Simultanübersetzung Deutsch - Englisch - Japanisch

Prof. Alex Waibel, Karlsruhe

JANUS ist ein Sprach-zu-Sprach-Übersetzungssystem. Es versteht kontinuierlich gesprochenes Deutsch und Englisch, ist unabhängig vom Sprecher, übersetzt ins Englische und Japanische und versteht das Konferenzanmeldungs-vokabular (500 Wörter).

Kontinuierliche Sprache wird über ein Mikrofon aufgenommen und von einem Analog/Digital Wandler digitalisiert. Von der Aufnahme werden alle zehn Millisekunden 16 spektrale Fourierkoeffizienten berechnet. Neuronale Netze und statistische Algorithmen berechnen aus den Koeffizienten Wahrscheinlichkeiten für Phoneme. Die Wahrscheinlichkeiten werden vom "Dynamischen Programmieren" verwendet, um eine oder mehrere Satzypothesen zu erstellen. Ein Parser, der auf LR-Grammatiken oder neuronalen Netzen basiert, analysiert die Hypothesen und übersetzt sie in die Zwischensprache Interlingua. Aus der Interlingua Repräsentation wird ein Satz in der Zielsprache generiert und über einen Sprachsynthesizer ausgegeben.

Das Sprach-zu-Sprach-Übersetzungssystem JANUS wird im Rahmen einer Kooperation zwischen der Universität Karlsruhe und der Carnegie-Mellon University in Pitts-

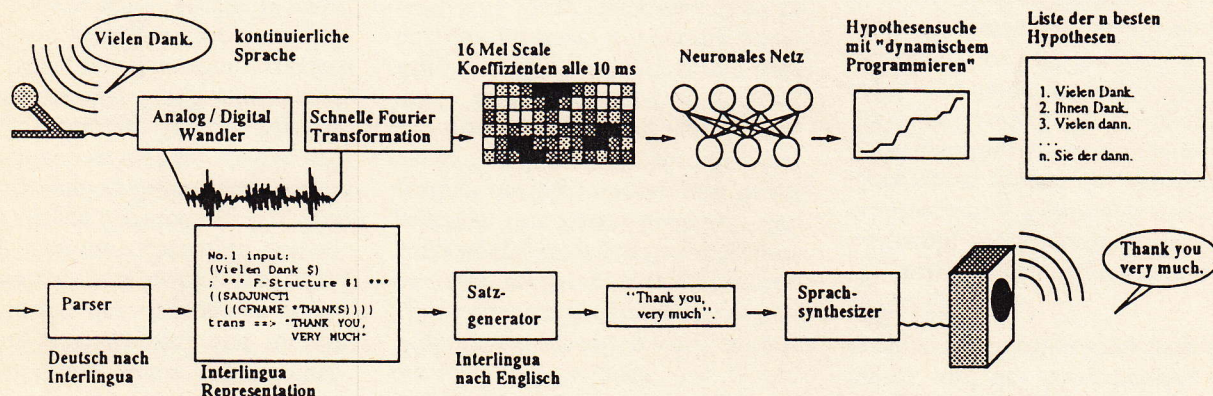
burgh, USA, unter der Leitung von Prof. Waibel entwickelt. JANUS erkennt von beliebigen Sprechern kontinuierlich gesprochene englische oder deutsche Sprache und übersetzt sie in die künstliche Zwischensprache "Interlingua". Aus Interlingua wird dann die Übersetzung in der Zielsprache generiert, die von einem Sprachsynthesizer ausgegeben wird.

Das System besteht aus mehreren unabhängigen Teilen, einem Spracherkenner, der Sprache in Text wandelt, einem Parser, der diesen Text analysiert und übersetzt, und einer Sprachausgabereinheit. Es hat einen modularen Aufbau, so daß die Zielsprache völlig unabhängig vom Spracherkenner ist. Jede Sprache kann als Zielsprache verwendet werden, sofern ein Satzgenerator existiert, der aus Interlingua Sätze in der Zielsprache erzeugt. Ebenso kann der Teil des Systems, der die Erkennung durchführt, ohne Einfluß auf die anderen Teile durch einen Erkenner für eine andere Sprache ersetzt werden. Auch der Parser, der den erkannten Text analysiert, kann durch einen anderen ersetzt werden. Als Zielsprachen unterstützt JANUS zur Zeit Englisch, Deutsch und Japanisch.

Das Erkennen der Sprache, also das Umwandeln von gesprochener Sprache in geschriebenen Text, geschieht mit Hilfe künstlicher neuronaler Netze und statistischer Methoden (Hidden Markov Modelle). Der Spracherkenner liefert eine Reihe möglicher Satzypothesen, von denen die beste aus der Eingabesprache nach Interlingua von speziellen LR-Parsern oder alternativ von künstlichen Neuronalen Netzen übersetzt wird.

JANUS ist zur Zeit noch aufgabenabhängig, das heißt, der verstandene Wortschatz ist beschränkt auf etwa 500 Wörter. Das System arbeitet in einem Szenario, in dem ein Anrufer bei einem Konferenzsekretariat anruft, um sich zu einer Konferenz anzumelden oder konferenzbezogene Fragen zu stellen.

JANUS braucht eine einmalige Trainingsphase, in der ihm Beispielsätze von vielen Sprechern präsentiert werden. Das vorhandene System wurde für die Erkennung englischer und deutscher Sprache mit je etwa 1000 Sätzen von etwa 60 verschiedenen Sprechern trainiert. Ist JANUS einmal trainiert, ist es einsatzbereit und muß nicht nachtrainiert werden. Ein neues Training ist erst erforderlich, wenn der zu erken-



Forschung und Lehre

nende Wortschatz erheblich größer wird oder wenn sich das Aufgabengebiet ändert. Das ausgestellte System verwendet für jede Übersetzungsrichtung einen Arbeitsplatzrechner (HP Apollo) und für die Ausführung der neuronalen Algorithmen einen hochparallelen SIMD-Rechner (MasPar MP-1). Außer diesen Geräten werden dann noch ein Analog/Digital Wandler für die Sprachaufnahme sowie ein Sprachsynthesizer für die Sprachausgabe benötigt. Der Benutzer muß in ein Nahsprechmikrofon sprechen. Mit dieser Konfiguration kann JANUS gesprochene Sprache beinahe in Echtzeit übersetzen.

Bei einem Übersetzungsvorgang wird das digitalisierte Sprachsignal mittels schneller diskreter Fouriertransformation vorverarbeitet. Die alle zehn Millisekunden erzeugten 16 Frequenzkoeffizienten werden den künstlichen neuronalen Netzen eingegeben, welche dann für jeden Zeitpunkt Wahrscheinlichkeiten dafür liefern, daß es sich beim Sprachsi-

gnal um ein bestimmtes akustisches Phänomen handelt. In der Regel sind Teile von Phonemen unter Berücksichtigung des phonetischen Kontextes solche Phänomene. Aus diesen Wahrscheinlichkeiten wird mittels "Dynamischen Programmierens" eine oder mehrere Satzhypothesen aufgestellt und an den Parser weitergegeben. Der Parser analysiert den Aufbau der Hypothesen und findet deren Sinn heraus. So wird das, was der Sprecher gemeint hat, sinngemäß in Interlingua festgehalten, woraus in der Zielsprache ein Satz mit der gleichen Bedeutung generiert wird. Dieser Satz wird dann an das intelligente Sprachsynthesegerät geschickt, welches ihn in akustische Signale wandelt und ausspricht. Ebenso wie zur Zeit noch alle Spracherkennungssysteme, hat auch JANUS Probleme, Sprache zu erkennen, die von Hintergrundgeräuschen gestört ist, sowie spontane Sprache, bei der der Benutzer gefüllte (Ähs) und ungefüllte Pausen machen und ungrammatikalische oder

abgebrochene Sätze sagen kann. Mit diesen Problemen beschäftigt sich die Forschung ebenso wie mit der Vergrößerung des erkannten Vokabulars und der Verarbeitung von qualitativ schlechten Sprachsignalen (z.B. Telefon).

In weiteren Projekten arbeitet der Lehrstuhl von Prof. Waibel an der multimodalen Kommunikation mit Computern. Neben der akustischen Spracherkennung wird auch an der On-Line Handschrifterkennung und am Lippenlesen geforscht. Bei der On-Line Handschrifterkennung soll es dem Benutzer ermöglicht werden, in seiner eigenen fließenden Handschrift ohne den Zwang, in Blockbuchstaben an bestimmten Stellen schreiben zu müssen, Eingaben in den Computer zu tätigen. Experimente mit Lippenlesen haben schon gezeigt, daß die Information, die von einer Kamera über den Zustand der Lippen eines Sprechers geliefert wird, vor allem dort helfen kann, wo der akustische Spracherkenner Schwierigkeiten hat. ●