

SpeechDat – eine große europäische Telefonsprachdatensammlung

Christoph Draxler
Institut für Phonetik und Sprachliche Kommunikation
Universität München
Schellingstr. 3
D 80799 München
Tel: +49/89/2866 9968
Email: draxler@phonetik.uni-muenchen.de

Zusammenfassung

SpeechDat ist ein von der EU gefördertes Projekt zur Sammlung von Telefonsprache nach einheitlichem Verfahren in 15 europäischen Sprachen. Ausdrückliches Ziel ist die Schaffung und allgemeine Bereitstellung von qualitativ hochwertigen Ressourcen für die gegenwärtigen und mittelfristigen Bedürfnisse der Sprachtechnologie und -forschung.

Der Inhalt der SpeechDat-Datenbasen umfaßt Ziffern, Zahlen, Geldbeträge, Datums- und Zeitausdrücke, Buchstabierungen, Eigen-, Firmen- und geographische Namen, phonetisch interessante Wörter und Sätze sowie Anweisungen zur Steuerung von Telefon- und Informationsdiensten (z.B. voice-dialling, Auskunftsdienste). Für das Telefonfestnetz, das Mobiltelefonnetz und die Sprecherverifikation gibt es getrennte Sprachdatenbasen.

SpeechDat-Aufnahmen erfolgen in digitaler Form auf ISDN-Sprachservern. SpeechDat-Teilnehmer rufen den Sprachserver an und beantworten in einem computergeführten Interview Fragen oder lesen aus einem Fragebogen vor. Die Sprachaufnahmen werden orthographisch verschriftet und mit Grobmarkierungen für Geräusche versehen.

Alle Datenbasen werden von einer unabhängigen Einrichtung formal und inhaltlich validiert und nach der Endabnahme über die European Language Resources Association (ELRA) bzw. den SpeechDat-Partner vertrieben.

Das SpeechDat-Projekt

Das SpeechDat-Projekt ist in zwei Phasen unterteilt: SpeechDat(M) von September 1994 bis Februar 1996 als Pilotprojekt, und SpeechDat(II) von Februar 1996 bis September 1998. In SpeechDat(M) entstanden 8 Sprachdatenbasen mit je 1000 Sprechern über das Telefonfestnetz in Dänisch, Deutsch, Englisch, Französisch (Frankreich und französische Schweiz), Italienisch, Portugiesisch, Spanisch sowie eine Sprachdatenbasis für das Mobiltelefonnetz mit 300 Sprechern in Italienisch.

In SpeechDat(II) werden drei Arten von Sprachdatenbasen erstellt: Festnetz-, Mobiltelefonnetz-, und Sprecherverifikationsdatenbasen. Über das Festnetz wird jeder Sprecher einmal aufgenommen, über das Mobiltelefonnetz entweder jeder Sprecher einmal, oder eine geringere Anzahl Sprecher mit jeweils mehreren Anrufen unter verschiedenen Bedingungen, z.B. zuhause und im Auto. Für die Sprecherverifikationsdatenbasen rufen Sprecher mindestens 20 mal über einen Zeitraum von einigen Monaten hinweg an (Tabelle 1).

Sobald ein Partner seine Datenbasis fertiggestellt hat, kann er sie projektintern verfügbar machen und hat damit ein Anrecht auf die Datenbasen der anderen Partner. Die Sprachdatenbasen mit mehr als 2000 Sprechern sind in eine früh abzuliefernde Sprachdatenbasis mit 1000 und eine zweite mit den restlichen Sprechern unterteilt. Diese Unterteilung erlaubt eine erste Entwicklung von Spracherkennern und sprachgesteuerten Anwendungen.

Sprache	Netz	Sprecher /Anrufe	Sprache	Netz	Sprecher /Anrufe
Dänisch	Festnetz	+ 1000/1	Griechisch	Festnetz	* 5000/1
	Festnetz	* 4000/1		Italienisch	Festnetz
Deutsch	Festnetz	+ 1000/1	Mobilnetz		+ 400/1
	Festnetz	* 4000/1	Festnetz		* 3000/1
	Mobilnetz	* 1000/1	Mobilnetz		* 250/4
Deutsch (Luxemburg)	Festnetz	* 500/1	Niederländisch	Mobilnetz	* 250/4
Deutsch (Schweiz)	Festnetz	* 1000/1	Niederländisch (Flämisch)	Festnetz	+ 1000/1
Englisch (Britisch)	Festnetz	+ 1000/1	Norwegisch	Festnetz	* 1000/1
	Festnetz	* 4000/1	Portugiesisch	Festnetz	+ 1000/1
	Mobilnetz	* 1000/1		Festnetz	* 4000/1
	Verifikation	* 120/20	Slowenisch	Festnetz	* 1000/1
Finnisch	Festnetz	* 4000/1	Spanisch	Festnetz	+ 1000/1
Französisch	Festnetz	+ 1000/1		Festnetz	* 4000/1
	Festnetz	* 5000/1	Schwedisch	Festnetz	* 1000/1
	Verifikation	* 120/20		Mobilnetz	* 1000/1
Französisch (Wallonisch)	Festnetz	* 1000/1	Schwedisch (Finnisch)	Festnetz	* 1000/1
Französisch (Luxemburg)	Festnetz	* 500/1	Walisisch	Festnetz	* 2000/1
Französisch (Schweiz)	Festnetz	+ 1000/1	* in Arbeit + fertiggestellt		
	Festnetz	* 2000/1			
	Verifikation	* 20/50			

Tabelle 1: SpeechDat Datenbasen

Inhalt

Die Definition des gemeinsamen Inhalts aller SpeechDat-Datenbasen ist in Tabelle 2 wiedergegeben.

Gegenüber SpeechDat(M) wurde dieser gemeinsame Inhalt um Eigen- und Firmennamen sowie geographische Bezeichnungen ergänzt; zudem wurde mehr Gewicht auf phonetisch reiches Material gelegt.

Die Anzahl der Aufnahmen der einzelnen Äußerungsklassen ist so bemessen, daß ein Training von Spracherkennern für die jeweilige Klasse möglich ist.

3/1	Befehlswörter und -ausdrücke	Wörter und Phrasen aus Telefoniediensten und Anwendungen
6	Ziffernfolgen	Fragebogen-, Kreditkarten, Geheim-, Telefonnummern, Ketten isoliert gesprochener Ziffern
2	Zahlen	isoliert gesprochene Ziffern und Zahlen
1	Geldbeträge	landesspezifische Währung
5	Datums- und Uhrzeitangaben	absolute und relative Angaben in numerischer und alphanumerischer Form
3	Buchstabierungen	Wörter, Eigennamen, Buchstabensequenzen
5	Auskunftsdienst	geographische Bezeichnungen, Eigen- und Firmennamen
2	ja/nein Antworten	spontane ja/nein Antworten auf Fragen
4/9	phonetisch reiche Wörter/Sätze	die Sätze eines Fragebogens enthalten jedes Phonem zweimal

Tabelle 2: SpeechDat(II) Inhalt

Demographische Anforderungen

Für SpeechDat(II) wurden die folgenden demographischen Anforderungen spezifiziert:

- je 50% ($\pm 2,5$ %) männliche und weibliche Sprecher
- jeweils mindestens 20% der Anrufe aus den drei Altersklassen 15-30, 31-45, 46-60 Jahre
- repräsentative regionale Abdeckung

Zusätzlich wurde vereinbart, möglichst auch Kinder und Jugendliche aufzunehmen.

Annotation

In SpeechDat werden die Anrufe orthographisch verschriftet. Numerische Angaben werden in die entsprechenden Zahlwörter umgewandelt und Buchstabierungen wortgetreu verschriftet.

- 2444 kann z.B. als „zwei tausend vier hundert vierundvierzig“, „vierundzwanzig vierundvierzig“ „zwo vier vier vier“, oder auch „zwei dreimal die vier“,
- IBM kann als „I B M“, „Ida Berta Martha“ oder auch „groß I groß B groß M“

geäußert werden.

Es werden mindestens vier Klassen von Geräuschen unterschieden und mit entsprechenden Marken in der Annotation gekennzeichnet: Hästitionsäußerungen wie „äh“, „ähm“, usw., andere Sprechergeräusche wie Husten, Lachen, usw., und nicht vom Sprecher herrührende kurze oder andauernde Geräusche wie Türschlagen, Berührungen des Telefonhörers, oder Musik und Gespräche im Hintergrund, Verkehrslärm.

Die Annotationen werden in separaten sogenannten Labeldateien im SAM-Format gespeichert. Labeldateien sind ISO 8859 Textdateien. Sie enthalten alle relevanten Angaben zu einer einzelnen Aufnahme.

LHD: SAM, 5.10
DBN: SpeechDat_German_Fixed_Network
VOL: FIXED1DE
SES: 0003
DIR: \FIXED1DE\BLOCK00\SES0003
SRC: A10003A1.DEA
CCD: A1
SHT: 4294-8
CMT: *** signal data ***
BEG: 0
END: 24000
REP: Dept. of Phonetics, University of Munich, Germany
RED: 14/Dec/1997
RET: 18:14:00
SAM: 8000
SNB: 1
SSB: 8
QNT: A-LAW
CMT: *** speaker data ***
SCD: UNKNOWN
SEX: F
AGE: 30
ACC: BY
CMT: *** environment data ***
REG: UNKNOWN
ENV: OFFICE
NET: PSTN
PHM: TOUCH-TONE
LBD:
CMT: *** transcription data ***
LBR: 0,24000,,,,Nachricht
LBO: 0,12000,24000,[spk] Nachricht
ELF:

Zeilen mit Label LBR: enthalten die Vorgabe des Fragebogens, Zeilen mit LBO: die orthographische Verschriftung der gesprochenen Äußerung.

Ein Aussprachelexikon enthält mindestens alle in den Verschriftungen vorkommenden Wörter mit optionaler Angabe der Frequenz und eine kanonische Aussprachevariante in SAM-PA Notation.

Abend 34 a: b @ n t

Validierung

Alle SpeechDat Sprachdatenbasen werden von einer unabhängigen Stelle validiert. In SpeechDat(II) ist das Centre for Speech Expertise (SPEX) in Nijmegen, NL, für diese Validierung zuständig.

Die Validierung erfolgt mehrstufig: vor Beginn der Aufnahmen in großem Umfang müssen alle SpeechDat-Partner eine kleine Test-Sprachdatenbasis von ca. 10 Sprechern im endgültigen SpeechDat-Format erstellen und diese auf formale Fehler hin prüfen lassen. Formale Fehler sind inkorrekte Dateisystemstrukturen und Dateinamen, die Verwendung nicht erlaubter Marken und Symbole, und ein mit den Verschriftungen inkonsistentes Lexikon.

Nach erfolgreicher Validierung der Testdatenbasis können die eigentlichen Aufnahmen beginnen. Nach Abschluß der Aufnahmen und der Verschriftung (bei den großen Datenbasen mit mehr als 2000 Sprechern sind die Aufnahmen in zwei Tranchen zu 1000 und die restlichen Sprecher unterteilt) wird die fertige Datenbasis erneut validiert. Bei dieser abschließenden Validierung wird auch eine Stichprobe der Verschriftungen kontrolliert. Eine Datenbasis wird nur dann abgenommen, wenn sie den SpeechDat-Spezifikationen genügt.

SpeechDat in Deutschland

Das Institut für Phonetik und Sprachliche Kommunikation (IPSK) der Universität München (Vorstand: Prof. Dr. H. G. Tillmann) führt die deutsche SpeechDat(II) Sprachdatensammlung für das

- Telefonfestnetz im Unterauftrag der SIEMENS AG, München, und das
- Mobiltelefonnetz im Unterauftrag der Vocalis Ltd, Cambridge,

durch.

Technische Installation

Ein 486-PC unter SCO Unix ist über eine aculab-Karte an einen Primärmultiplex ISDN Anschluß der Deutschen Telekom angeschlossen. Die Software zur Dialogsteuerung und der Aufnahme des Sprachsignals wurde am Institut entwickelt. Sie erlaubt die gleichzeitige Aufnahme von max. 6 Kanälen. Der Server steht rund um die Uhr zur Verfügung; aus Sicherheitsgründen werden alle Aufnahmen sowohl auf lokalen Rechnern gespiegelt als auch am Leibniz-Rechenzentrum gesichert. Diese Sicherung erfolgt ohne Unterbrechung des Serverbetriebs in den verkehrsarmen Nachtstunden.

Rekrutierung

Die Rekrutierung der Sprecher in Deutschland erfolgte auf drei verschiedene Vorgehensweisen:

- siemensinterne hierarchische Rekrutierung: Abteilungs- oder Gruppenleiter wurden gebeten, Fragebogen an mindestens 10 Mitarbeiter zu verteilen.
- Aufruf zur Teilnahme an SpeechDat: in der Tagespresse und der Siemens Hauszeitschrift SiemensWelt wurde das Projekt vorgestellt und eine Telefon- bzw. Faxnummer zur Anforderung von Fragebögen veröffentlicht.
- Werbung im Internet: auf den WWW-Seiten von SpeechDat befinden sich online-Fragebögen bzw. Verweise auf die Datensammlungen in den einzelnen Ländern.

Die siemensinterne Rekrutierung erbrachte einen Rücklauf von unter 10%. Veröffentlichungen in der Tagespresse ergaben maximal 60 Fragebogenanforderungen, von denen dann ca. 60% tatsächlich zu Anrufen führten; die Aufrufe in der SiemensWelt und den vdi-nachrichten war mit jeweils mehr als 200 Fragebogenanforderungen die erfolgreichsten Aufrufe. Die Werbung im Internet führte bislang zu insgesamt ca. 90 Anrufen.

Verschriftung

Zur Verschriftung der Aufnahmen wird am IPSK das System WWWTranscribe (Abb. 1) eingesetzt. WWWTranscribe basiert auf dem WWW, d.h. es ist plattformunabhängig und ermöglicht die Verschriftung von jedem Internetzugang aus oder auch lokal von CD-ROM.

WWWTranscribe unterstützt die Verschriftung durch spezielle Editierhilfen wie z.B. Buttons zur Konversion von numerischem Format in die entsprechende orthographische Form oder eine Syntaxkontrolle der Verschriftung.



Abbildung 1: WWWTranscribe

Stand der Aufnahmen und Verschriftungen

Im September 1997 waren zwei wichtige Meilensteine erreicht:

- die deutsche SpeechDat(II) Datenbasis mit den ersten 1000 Sprechern wurde validiert und abgenommen, und
- mit 2000 Anrufen war die Hälfte der aufzunehmenden Anrufe erreicht.

Acht studentische Hilfskräfte mit phonetischem Wissen sind zur Zeit für die Verschriftung neuer Anrufe und die Stichprobenkontrolle von Verschriftungen eingesetzt. Die Verschriftung eines Anrufs von ca. 4 Minuten Sprache dauert insgesamt ca. 40 Minuten. Bei der Stichprobenkontrolle werden einzelne Aufnahmen zufällig ausgewählt und gegebenenfalls korrigiert.

Die Validierung der ersten 1000 Sprecher durch das SPEX war insgesamt erfolgreich. Einige Kriterien wurden nicht erfüllt: die Anzahl der Sprecher ist zu gering, einzelne Äußerungsklassen enthalten eine zu geringe Anzahl Ausdrücke, und die Signalanalyse fehlt. Die Fehlerquote bei der Verschriftung ist in Tabelle 3 wiedergegeben.

	Sprache	Marken
kurze Items	2,1 %	5,7 %
lange Items	6,3 %	5,9 %

Tabelle 3: Fehlerquoten der Validierung der ersten 1000 Sprecher

Kurze Items sind Ziffern, Datumsausdrücke, Namen und Wörter, lange Items sind die Sätze, Buchstabierungen, Ziffernketten u.ä.

Diese Datenbasis enthielt nur 988 statt 1000 Sprecher, da zwei doppelt vorhandene Sprecher unmittelbar vor dem Brennen CD-ROMs entfernt wurden. Bei den Städtenamen und den PINs wurden weniger als die angegebenen 500 bzw. 150 unterschiedliche Ausdrücke aufgenommen, da keine Fragebogen mit diesen Angaben beantwortet wurden bzw. die Menge unterschiedlicher Fragebögen geringer war als die Anzahl Ausdrücke in der Klasse. Die Signalanalyse fehlt, weil die Software zum Zeitpunkt der Validierung nicht lauffähig war.

Im Januar 1998 wurden für das

- Festnetz 3600 Anrufe, und für das
- Mobilnetz 320 Anrufe

aufgenommen. Davon wurden bis dahin 3050 bzw. 70 verschriftet.

SpeechDat im WWW

Das IPSK unterhält den zentralen WWW Server für SpeechDat(II):



<http://speechdat.phonetik.uni-muenchen.de/>

Der größte Teil des Servers ist frei zugänglich. Sämtliche SpeechDat(II) Spezifikationsdokumente und Standards sind öffentlich verfügbar, ebenso Beispiele von Signaldateien und Annotationen in den verschiedenen Sprachen.

Informationsmaterial

Auf dem SpeechDat(II) WWW Server sind die Flaggen auf der Startseite mit den WWW Seiten der SpeechDat-Partner verbunden. Die deutsche SpeechDat-Seite enthält einige der bisher veröffentlichten SpeechDat-Aufrufe in der Tagespresse und einen online Fragebogen.

Unter dem Stichwort „Deliverables“ befinden sich eine Übersicht über Inhalt und Umfang aller SpeechDat-Datenbasen, sowie SpeechDat-Poster.

Aufruf zur Teilnahme

Das IPSK sucht Personen mit Muttersprache Deutsch, die den SpeechDat-Sprachserver anrufen. Der Anruf ist gebührenfrei. Jeder Anrufer, der ein ausgefülltes Datenblatt zurückschickt, erhält für einen Anruf im Festnetz als Belohnung eine Telefonkarte im Wert von 6.- DM. Zudem nimmt er an einer Verlosung von DM 1000.- am 27.März 1998 teil.

Fordern Sie Ihren Fragebogen an unter

Tel: 089/286 6940

Fax: 089/280 0362

oder im WWW

<http://speechdat.phonetik.uni-muenchen.de/>

unter der deutschen Flagge.

Literatur

Chr. Draxler

WWWTranscribe – A Modular Transcription System Based on the WWW; Eurospeech '97, Rhodos, 1997

K. Kordi

Definition of Corpus, Scripts, and Standards for Speaker Verification, SpeechDat Report LE2-4001-SD1.1.3, 1996

F. Senia

Specification of Speech Database Interchange Format (Version 4.4), SpeechDat Report LE2-4001-SD1.3.1, 1997

M. Tomlinson, R. Winski, W. Barry

Label file format proposal, Esprit project 1542 (SAM): Extension Phase, Final Report, 1988

H. van den Heuvel

Validation Criteria, SpeechDat Report LE2-4001-SD1.3.3

J. G. van Velden, D. Langmann, M. Pawlewski

Specification of Speech Data Collection over Mobile Telephone Networks, SpeechDat Report LE2-4001-SD1.1.2/1.2.2, 1996

R. Winski

Definition of corpus, scripts and standards for Fixed Networks, SpeechDat Report LE2-4001-SD1.1.1, 1997