

Aus dem Institut für Röntgendiagnostik  
der Medizinischen Fakultät Charité  
der Humboldt-Universität zu Berlin  
Direktor: Prof. Dr. med. B. Hamm



# Dissertation

## Einsetzbarkeit und Nutzen der digitalen Spracherkennung in der radiologischen Diagnostik

zur Erlangung des akademischen Grades  
doctor medicinae  
(Dr. med.)

vorgelegt der Medizinischen Fakultät Charité  
der Humboldt-Universität zu Berlin

von

Herrn **Holger Arndt**

aus: Berlin

Dekan: Prof. Dr. med. M. Dietel

Gutachter: Prof. Dr. med. B. Hamm  
PD Dr. med. S. Mutze  
.Prof. Dr. med. K.-J. Wolf

Datum der Promotion: 17.2.1999

# Inhaltsverzeichnis

<b>Verzeichnis der verwendeten Abkürzungen / Begriffe.....</b>	<b>5</b>
<b>1 Einleitung .....</b>	<b>6</b>
<b>1.1 Allgemeine Forderungen an die radiologische Befundung .....</b>	<b>6</b>
<b>1.2 Grundlagen der digitalen Spracherkennung .....</b>	<b>9</b>
1.2.1 Entwicklung der digitalen Spracherkennung.....	9
1.2.2 Grundlagen für die Kommunikation Mensch/Maschine .....	10
1.2.3 Computer und Mikroelektronik.....	11
1.2.4 Sprachanalyse und Sprachsynthese .....	12
1.2.5 Spracherkennung und Sprachverständnis .....	15
1.2.6 Stand der Entwicklung und Forschung in der digitalen Sprachverarbeitung .....	17
<b>1.3 Einsatz der digitalen Spracherkennung in der Medizin.....</b>	<b>18</b>
<b>2 Zielstellung.....</b>	<b>20</b>
<b>3 Material und Methoden.....</b>	<b>20</b>
<b>3.1 Das Spracherkennungssystem SP 6000 .....</b>	<b>20</b>
3.1.1 Diktat .....	21
3.1.2 Spracherkennung .....	22
3.1.3 Korrektur.....	22
3.1.4 Übergabe des Befundtextes.....	23
3.1.5 Lexikon .....	23
3.1.6 Initialtraining und Adaptation .....	24
3.1.7 Konfiguration des Systems.....	25
<b>3.2 Testpersonen .....</b>	<b>25</b>
<b>3.3 Erfasste Werte .....</b>	<b>26</b>
<b>3.4 Datenmenge .....</b>	<b>27</b>
<b>3.5 Statistische Auswertung.....</b>	<b>29</b>
3.5.1 Mittelwert .....	29
3.5.2 Standardabweichung.....	29
3.5.3 Ungepaarter Wilcoxon-Test (Mann-Whitney U-Test).....	30
<b>4 Ergebnisse .....</b>	<b>31</b>
<b>4.1 Arbeitsaufwand zur Installation, Netzwerkeinbindung und Kopplung des Systems an das bestehende RIS.....</b>	<b>31</b>
<b>4.2 Zeitaufwand für das Initialtraining .....</b>	<b>31</b>
<b>4.3 Zeitaufwand für die Betreuung des laufenden Systems .....</b>	<b>32</b>
<b>4.4 Entwicklung der Fehlerrate .....</b>	<b>32</b>
4.4.1 Fehlerrate nach dem Initialtraining .....	32
4.4.2 Fehlerrate nach der ersten Adaptation .....	35
4.4.3 Fehlerrate nach der zweiten Adaptation .....	36
4.4.4 Weitere Entwicklung der Fehlerrate (nach der dritten Adaptation) .....	38

4.4.5 Fehlerrate in Abhängigkeit vom Individuum.....	40
4.4.6 Fehlerrate in Abhängigkeit von der Untersuchungsart und -anzahl.....	40
4.4.7 Fehlerrate in Abhängigkeit von der Diktatdauer .....	42
<b>4.5 Kontrolle und Korrektur in Abhängigkeit von der Fehlerrate .....</b>	<b>43</b>
<b>4.6 Beurteilung der Befundungszeit.....</b>	<b>44</b>
4.6.1 Zeitdauer zur Befunderstellung mittels SP 6000 .....	44
4.6.2 Zeitdauer zur Befunderstellung mittels Schreiben .....	46
4.6.3 Vergleich der Befundungszeit des Schreibens mit der des Diktierens .....	48
<b>5 Diskussion.....</b>	<b>50</b>
<b>5.1 Methodik.....</b>	<b>50</b>
5.1.1 Testpersonen .....	50
5.1.2 Meßwerte.....	50
<b>5.2 Entwicklung der Fehlerrate .....</b>	<b>51</b>
<b>5.3 Kontrolle und Korrektur .....</b>	<b>53</b>
<b>5.4 Befundungszeit.....</b>	<b>55</b>
<b>6 Zusammenfassung .....</b>	<b>58</b>
<b>7 Literatur .....</b>	<b>62</b>
Eidesstattliche Erklärung.....	65
Danksagung.....	66
Lebenslauf.....	67

## Verzeichnis der verwendeten Abkürzungen / Begriffe

A-D-Wandlung	Analog-Digital-Wandlung
Angio	Untersuchungen der Gefäße
CT	Computertomographie
D-A-Wandlung	Digital-Analog-Wandlung
DOS	disk operating system - Betriebssystem für Personalcomputer
FKDS	Farbkodierte Dopplersonographie
IC-	integrated circuit / Mikroelektronik
HNO	Untersuchungen des Hals-Nasen-Ohrenbereichs
IP-Adresse	eindeutige Identifizierung eines Computers im Netzwerk
MRT	Magnetresonanztomographie
Novel Netware	Netzwerksoftware
NTX	Nierentransplantate
PC	Personal Computer
RIS	Radiologisches Informationssystem
RTF	Rich-Text-File - allgemeines Dateiformat für formatierte Texte
TCP/IP	Netzwerktransportprotokoll
US	Ultraschall
VOM	Vokabelmanager
ZVK	Zentraler Venenkatheter

## 1 Einleitung

### **1.1 Allgemeine Forderungen an die radiologische Befundung**

An die bildgebende Diagnostik werden zwei grundlegende Forderungen gestellt. Zum einen ist eine qualifizierte Befundung, die klinische Daten, Hergang bzw. Art der Diagnostik, Befundbeschreibung und die radiologische Diagnose enthält, gefordert. Zum anderen erwartet der Kliniker, daß die Befundung in einem adäquaten Zeitraum erfolgt und ihm mitgeteilt wird. Die maximal zu tolerierende Zeitspanne wird durch die klinische Fragestellung bestimmt. Während z.B. bei der Verlaufskontrolle einer konsolidierten Fraktur in der konventionellen Skelettdiagnostik durchaus ein bis zwei Tage zwischen der Untersuchung und dem Versand des Befundes an den Kliniker toleriert werden können, müssen die Ergebnisse anderer Untersuchungen, wie z.B. einer ZVK-Kontrolle, sofort vorliegen. Im letzteren Fall sollte der akute Befund natürlich als erstes telefonisch übermittelt werden, aber auch die nachfolgende schriftliche Befundung muß schnellstens erfolgen. Sicher kann im Klinikbereich bei der Festlegung der Priorität der Befundung auch in stationäre und ambulante Patienten unterteilt werden. Da die ambulant betreuten Patienten in der Regel erst nach einigen Tagen wieder in die poliklinische Sprechstunde bestellt werden, ist hier der zeitliche Abstand zwischen der eigentlichen Untersuchung und der schriftlichen Beurteilung variabler als bei stationären Patienten. Hier werden zumeist keine langfristigen Verlaufskontrollen durchgeführt, vielmehr entscheidet die bildgebende Diagnostik im Zusammenhang mit der Klinik über die aktuell durchzuführende Therapie, wie z.B. bei der Beurteilung der Progredienz oder Regredienz pneumonischer Infiltrationen als Anhalt für die Effektivität der antibiotischen Therapie oder das Staging von Tumoren zur Beurteilung der Operabilität. Aber auch diese Einteilung in akut und weniger akut zu befundende Untersuchungen birgt ein großes Risiko. Bei allen nicht schleunigst betrachteten und befundeten radiologischen Untersuchungen (welche z.B. als ambulant und nicht dringlich eingestuft wurden) besteht die potentielle Möglichkeit einen Zufallsbefund zu spät festzustellen bzw. an den klinischen Partner zu übermitteln und damit Zeit für eine spezielle Diagnostik oder schnelle Therapie zu verschenken.

Das heißt, es können keine Abstriche von den grundlegenden Anforderungen an die bildgebende Diagnostik gemacht werden.

Nach einer eventuell telefonischen Übermittlung bei einer akuten Diagnose ist der schriftliche Befund der durchgeführten Untersuchung sehr wichtig. Dem klinischen Kollegen gibt er die rechtliche Sicherheit eines Dokuments, anhand dessen er die Therapie durchführen

kann. Desweiteren kann der Kliniker auf Grundlage der schriftlichen Beurteilung besser mit Kollegen diskutieren, als wenn er ein Telefonat zitieren würde. Für den Radiologen wiederum ist der schriftliche Befund im Zusammenhang mit den angefertigten Aufnahmen ein Nachweis über die erbrachten Leistungen und deren Qualität. Ferner bietet der ausführliche, schriftlich fixierte Befund gerade bei langfristigen Verlaufskontrollen einen Anhaltspunkt für eine Verlaufsbeurteilung, auch wenn die Voraufnahmen gerade nicht verfügbar sind.

Welche Möglichkeiten gibt es für den Radiologen einen schriftlichen Befund zu erstellen?

Als schnellste und einfache Erstellungsweise wäre der handgeschriebene Befund zu nennen. Dieser ist jedoch im Zeitalter der elektronischen Daten- und Leistungserfassung obsolet und allenfalls für eine provisorische Befundübermittlung verwendbar.

Ein ordentlicher Befund muß heutzutage mit einem Computer geschrieben und in einem Radiologischen Informationssystem abgelegt werden. Nur so läßt sich eine exakte Leistungserfassung erstellen. Gleichzeitig ermöglicht das Informationssystem die Ansicht früherer Befunde ohne das die vollständige Patientenakte zur Verfügung stehen muß. Wie läßt sich nun ein computergeschriebener Befund erstellen?

Der Radiologe könnte den Befund selber am Computer schreiben. Diese Vorgehensweise hat verschiedene *Vorteile* :

- Nach der Beendigung der Befundung liegt der schriftliche Befund sofort vor.
- Der Radiologe hat in den meisten zur Verfügung stehenden RIS bzw. Textverarbeitungssystemen die Möglichkeit, Textbausteine für häufig vorkommende Beschreibungen oder komplette Befunde einzuarbeiten und über Tastaturkombinationen abzurufen.
- Bei kurzfristigen Verlaufsbeurteilungen können die im RIS vorliegenden Vorbefunde als Ausgang für die aktuelle Befundung genutzt werden.
- Die Möglichkeit der Nutzung von vorgegebenen Textbausteinen und Vorbefunden erleichtert dem unerfahrenen Kollegen den eigenen Befundungswortschatz aufzubauen bzw. zu erweitern.
- Bis zur Befundausgabe (speichern, drucken und absenden) kann der Befundtext direkt verändert werden.

Durch das Schreiben entstehen aber auch *Nachteile*:

- Es bestehen große interindividuelle Unterschiede in der Tastaturfertigkeit des Einzelnen. Während wenige im Zehnfingersystem mit einer hohen Geschwindigkeit schreiben können, benutzen die meisten Kollegen die Tastatur im „Ein-Finger-Suchsystem“. Das bedeutet eine starke Belastung für die ungeübten Tastaturnutzer sowie eine verlängerte Befunderstellung.

- Da viele Radiologen nicht ohne Textkontrolle schreiben können, wandert der Blick ständig von der Untersuchung am Lichtkasten zu der Tastatur und zum Bildschirm, was wiederum eine erhöhte Belastung für den Befunder bedeutet, da er sich ständig aufs Neue in das vor ihm hängende Bild vertiefen muß. Dazu kommt noch die Belastung der Augen, da häufig der Fokus zwischen Augen-Monitor- und Augen-Bildabstand wechselt.
- Die Hände sind ständig an die Tastatur gebunden. Schon das Auf- oder Abhängen von Aufnahmen, um sie z.B. mit einer Lupe zu betrachten, erfordert einen erhöhten Aufwand.
- Während des Schreibens ist der Radiologe an den Schreibplatz gebunden (meist in sitzender Position), was bei Untersuchungen mit mehreren Aufnahmen (z.B. MRT oder CT) die Beurteilung erschwert, da der Befunder, um alle Details erkennen zu können, den Schreibplatz oft verlassen muß.

Eine andere Möglichkeit der schriftlichen Befunderstellung bietet das Diktieren der Untersuchungsergebnisse. Nach Beendigung der Befundung wird das Diktat an eine Schreibkraft übergeben, die dann den gesprochenen Text in den Computer eintippt.

Gegenüber der vorherigen Methode bieten sich verschiedene *Vorteile*:

- Die Beurteilung der Untersuchung ist nicht an einen bestimmten Arbeitsplatz gebunden.
- Die Aufmerksamkeit des Untersuchers kann vollständig auf die Aufnahmen konzentriert werden.
- Zumindest eine Hand ist frei (die andere bedient das Diktiergerät).
- Die Zeitdauer der Befundung hängt nicht von der Tastaturfertigkeit des Untersuchers ab.

Aber auch das Diktat bietet *Nachteile*:

- Im Regelfall erfolgt das Diktat mit einem Diktiergerät auf eine Kassette, die zu der Schreibkraft transportiert werden muß. Oft werden mehrere Diktate auf eine Kassette gesprochen, so daß das erste Diktat erst geschrieben werden kann, wenn alles diktiert bzw. die Kassette voll ist, wodurch eine Verzögerung der Befundschreibung entsteht.
- Die Zeit bis zur Fertigstellung des Befundes hängt zum einen von den Fähigkeiten und der Belastbarkeit der Schreibkraft und zum anderen von der Menge der anfallenden Diktate ab.
- In dem schriftlichen Befund können Fehler infolge Tippfehlern oder akustischen Verständigungsschwierigkeiten auftreten. Daher ist es notwendig, daß der Radiologe den geschriebenen Text auf fachliche und andere Fehler kontrolliert und gegebenenfalls korrigieren läßt.

Der Transport der Diktate über eine Kassette kann durch ein vernetztes Sprachsystem ersetzt werden. Bei einem derartigen System wird das Diktat digitalisiert und zentral gespei-

chert und steht damit sofort nach Beendigung des gesprochenen Diktats den Schreibkräften zur Verfügung. Letztendlich werden die besseren Bedingungen bei der Beurteilung einer Untersuchung beim Diktieren trotz allem durch einen verlängerten Zeitraum zwischen Beurteilung der Untersuchung und dem Versand des schriftlichen Befundes erkauft. Beim individuellen Schreiben wiederum liegt der Befund sofort vor, es besteht aber ein erhöhter Arbeitsaufwand für den Radiologen. Eine ähnliche Problematik stellt sich auch in anderen medizinischen Bereichen (20, 29, 36, 46, 48, 49).

Als interessante Alternative werden seit einiger Zeit digitale Spracherkennungssysteme zur Umwandlung von gesprochenem in geschriebenen Text angeboten. Diese Systeme verarbeiten das Sprachsignal und erstellen nach verschiedenen Algorithmen den geschriebenen Text (44).

## **1.2 Grundlagen der digitalen Spracherkennung**

### 1.2.1 Entwicklung der digitalen Spracherkennung

Es gibt viele Möglichkeiten der Kommunikation der Menschen untereinander, wie z.B. Gestik, Bilder, Zeichnungen, gedruckter Text und die Sprache. Im alltäglichen Leben setzte sich hauptsächlich die Möglichkeit der sprachlichen Verständigung durch. Die Gründe dafür wurden von Flanagan (10) zusammengefaßt: „Ich kann mir vorstellen, daß der Antrieb für die Sprachentwicklung nicht so sehr die Notwendigkeit des Ausdrucks der eigenen Gedanken war (was sicher über die Gestikulation ausreichend ermöglicht wurde), als vielmehr die Schwierigkeit mit ‘vollen Händen zu sprechen’.“

Letztendlich ist die Sprache der einzige effektive Weg für den Menschen seine Gedanken und Wünsche auszudrücken. Da ist es nicht verwunderlich, daß schon immer der Wunsch bestand, sprachlich mit Maschinen zu kommunizieren bzw. sie zu steuern.

Als in der menschlichen Geschichte Tiere zum Antrieb von Maschinen genutzt wurden, bot sich die Möglichkeit, über eine Kommandosprache die Tiere und damit die Maschinen zu lenken. Natürlich wurden die Tiere auch über Hilfsmittel wie z.B. Zaumzeug gesteuert, die Sprachsteuerung ermöglichte jedoch eine anderweitige Nutzung der Hände. Zum Beispiel war es möglich, mit einem Ochsenpflug ein Feld zu pflügen und gleichzeitig mit den Händen die Saat auszubringen bzw. Steine aufzusammeln. Diese limitierte Sprachsteuerung verschwand, als Kraftmaschinen (Dampfmaschine, Benzinmotor,...) nach und nach die Tiere als Antrieb verdrängten.

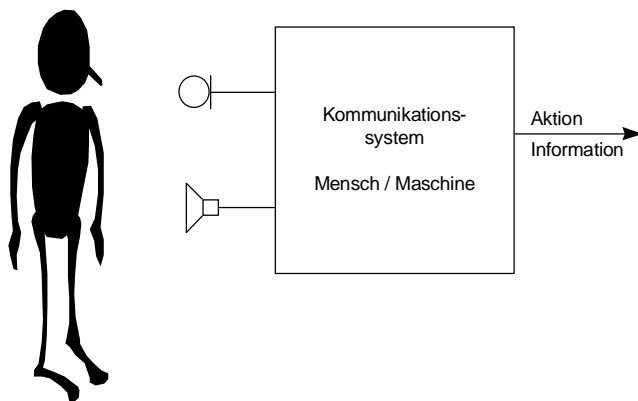
Bei immer komplexeren technischen Abläufen wird zum einen die Notwendigkeit einer Sprachsteuerung zur vereinfachten Kommunikation mit der Technik und zum anderen die Anforderung an die Zuverlässigkeit einer Sprachsteuerung deutlich.

Vorteile der sprachlichen Kommunikation mit Maschinen:

- Sprache ist für den Menschen die natürlichste Art und Weise der Kommunikation.
- Die Sprache ist auch verfügbar, wenn Hände und Augen anderweitig beschäftigt sind.
- Eine Sprachsteuerung von Maschinen wäre eine große Hilfe im alltäglichen Leben für Behinderte.
- Die Verknüpfung von Sprachsynthese, Spracherkennung und Sprachverständnis würde die vereinfachte Kommunikation zwischen verschiedenen Landessprachen ermöglichen.

### 1.2.2 Grundlagen für die Kommunikation Mensch/Maschine

Ein Kommunikationssystem Mensch/Maschine (Abb. 1) benötigt ein Mikrophon, um die



**Abb. 1 : Kommunikationssystem Mensch/Maschine**

menschliche Stimme aufzunehmen, sowie einen Lautsprecher oder Kopfhörer, um entweder eine Kontrolle des gesprochenen Textes zu ermöglichen oder dem Benutzer über eine synthetische oder vorher aufgenommene Stimme Reaktionen auf den gesprochenen Text mitzuteilen (44). So hat der Mensch die Möglichkeit mit dem System sprachlich zu kommunizieren. Das Sprachsystem muß den gesprochenen Text in der jeweiligen Art und Weise auswerten und die Informationen

weiterleiten. Es besteht die Möglichkeit, den gesprochenen Text im Sinne von Kommandos zu Steuerungsvorgängen zu nutzen, den Text zu erkennen und als geschriebenen Text bzw. symbolisch darzustellen. Beispiele für das Ausführen von sprachlichen Kommandos wäre ein System, das Schaltfunktionen ausführt, die Steuerung von Fahrzeugen bzw. einzelner Funktionen des Fahrzeugs. Ein anderes Anwendungsgebiet wäre der Zugriff und die Steuerung von Informationen. Das System könnte auf einen sprachlichen Auftrag hin in einer Datenbank suchen, eine Berechnung erstellen und das Ergebnis auch in sprachlicher Form ausgeben.

Im Vordergrund des Kommunikationssystems steht sicher an erster Stelle das Problem der Spracherkennung bzw. des Sprachverständnisses. Jedoch auch die Sprachsynthese ist gerade bei komplexeren Systemen, z.B. als Nachricht über die korrekte Erkennung und Verarbeitung des gesprochenen Textes oder als Antwort auf eine Informationsabfrage ein wichtiger Bestandteil. Nicht zu vergessen ist die Notwendigkeit der einfachen Bedienung durch den Nutzer.

Die Fragen, die Pierce (38) schon 1969 aufwarf, beschäftigten sich mit dem Ziel, dem Wert und dem potentiellen Erfolg der Forschung im Gebiet der Spracherkennung. Er stimulierte damit wesentlich die Diskussion, die Gedanken und den Enthusiasmus der Wissenschaftler und Ingenieure in den späten 60-er Jahren (44). In den letzten 25 Jahren wurden große Fortschritte in beiden Gebieten, der Spracherkennung und der Sprachsynthese gemacht, so daß bereits heute komplexe Sprachkommunikationssysteme möglich sind und auch in einzelnen Applikationen zur Verfügung stehen. Um die Entwicklung der Spracherkennung/ -synthese weiter voranzutreiben, ist eine enge Zusammenarbeit verschiedener wissenschaftlicher und technischer Gebiete, wie Psychologie, Linguistik, Akustik, Signalverarbeitung, Computertechnik sowie die IC-Technik notwendig.

### 1.2.3 Computer und Mikroelektronik

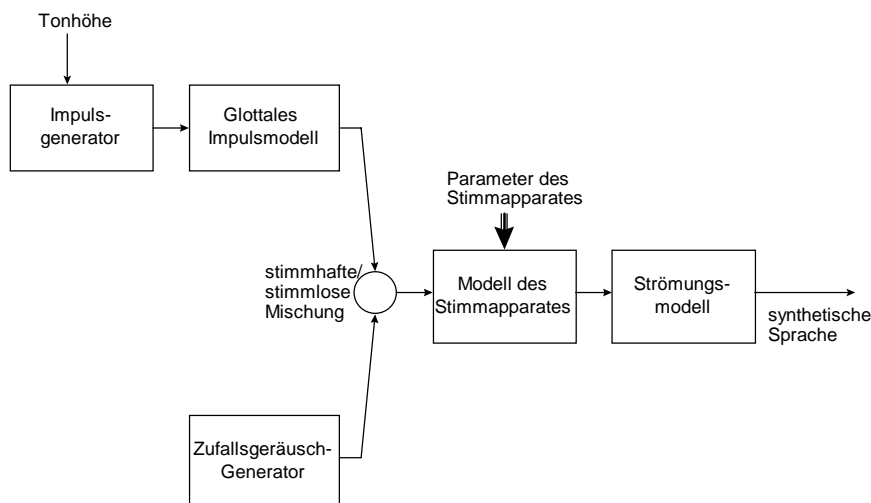
Seit über hundert Jahren erforschen Wissenschaftler und Ingenieure systematisch das Sprachsignal und die sprachliche Kommunikation. Die größten Fortschritte wurden aber erst durch den Einsatz digitaler Computer erzielt. Zuerst nutzte man die Rechner zur Simulation analoger Systeme. Dabei stellte sich heraus, daß komplexe Sprachverarbeitung nur mittels Computer möglich war. Ein großer Faktor für die beschleunigte Entwicklung war die Entwicklung digitaler Filter, der Spektrumanalyse sowie signalverarbeitender Techniken zur Sprachanalyse und Sprachsynthese (13, 37, 39). Außerdem stellt die programmierbare „Intelligenz“ der Computer einen Grundbaustein für die sprachliche Kommunikation zwischen Mensch und Maschine dar. Ohne die fulminante Entwicklung der Mikroelektronik wäre die Entwicklung von sprachgesteuerten Kommunikationssystemen nur im Labormaßstab denkbar. Innerhalb weniger Jahre hat sich sowohl die Dichte der Transistoren auf einem einzelnen Chip als auch die Taktgeschwindigkeit vervielfacht (44). Zusätzlich wurde durch multiple bzw. parallele Bustechnologien in der Schaltungsarchitektur eine hohe Verarbeitungsgeschwindigkeit erreicht, so daß schon auf herkömmlichen PC's Operationen für die Verarbeitung der anfallenden Daten bei einem Kommunikationssystem in Echtzeit ablaufen können. Die weitere Nutzung der Computer und der Mikroelektronik wird sich in Ab-

hängigkeit von der weiteren Entwicklung der IC-Technologien, der Erarbeitung effizienterer Algorithmen zur Signalverarbeitung sowie dem Einsatz der Multiprozessortechnik entfalten.

### 1.2.4 Sprachanalyse und Sprachsynthese

Bei einem Kommunikationssystem Mensch/Maschine ist die Sprache das informationenenthaltende Medium. Daher ist es notwendig, die Grundlagen des Sprachsignals zu kennen. Wie wird es erzeugt, wie wird die Information darin codiert und wie wird es empfangen?

Die menschliche Sprache ist eine Schallwelle, die durch ein komplexes physikalisches System erzeugt wird. Es ist möglich, den Vorgang der Spracherzeugung über physikalische Gesetze zu veranschaulichen bzw. zu simulieren. Seit vielen Jahren beschäftigen intensive Forschungen mit diesem Problem, deren Ergebnisse in klassischen Monographien (9, 10) sowie auch in neueren Arbeiten (39, 6) zu finden sind.

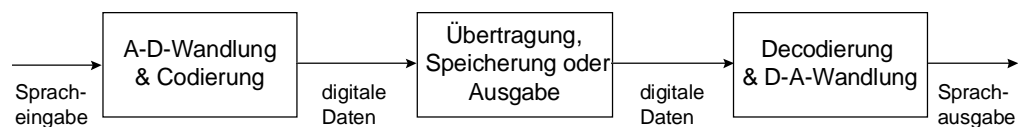


**Abb. 2 : Modell der Spracherzeugung - source system model (44)**

Die Forschung orientiert sich dabei hauptsächlich an einem klassischen Modell der Spracherzeugung, dem sogenannten 'source system model' (Abb. 2). Bei diesem Modell wird die Sprache über einen Wechsel der Anregung durch periodische Impulse (Tonhöhe) für stimmliche Klänge und ein Zufallsgeräusch für nichtstimmliche Klänge über eine Mischung beider Geräuschquellen erzeugt. Der Stimmapparat gestaltet zeitabhängig das gegebene Signal um, um geeignete Resonanzen oder Formanten zu erzeugen. Die Sprachanalyse beschäftigt sich mit der Ermittlung der Parameter des Modells, während die

Sprachsynthese sich mit der Ausgabe des Modells bei vorgegebenen Parametern befaßt (44).

Ein zentrales Problem der Sprachverarbeitung ist das Umwandeln der analogen Schallwellen in digitale Daten. Sprache, wie auch jede andere bandbreitenbegrenzte Wellenart, kann mittels einem Analog-Digital-Wandler in eine Abfolge binärer Daten umgewandelt werden. Diese digitalen Daten repräsentieren das Sprachsignal und können mittels Digital-Analog-Wandler wieder in ein analoges Signal umgewandelt und so wieder ausgegeben werden (Abb. 3).



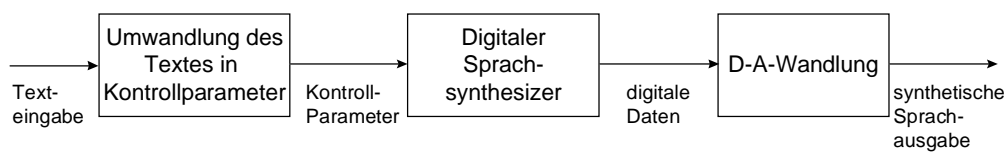
**Abb. 3 : Wandlung, Codierung, Decodierung**

Der digitale Wert, den der A-D-Wandler zu einem bestimmten Zeitpunkt erzeugt, wird auch als Sample bezeichnet. Je kürzer der Abstand zwischen der Ermittlung der einzelnen Sample-Werte ist, desto genauer wird die Änderung des analogen Signals erfaßt. Die Bitbreite bestimmt die Werte, die ein Sample annehmen kann. Bei einer Bitbreite von 4 Bit pro Sample könnte ein ganzzahliger Wert von 0 bis 15 angenommen werden. Eine Bitbreite von 16 Bit ermöglicht einen Sample-Wert von 0 bis 65535. Das heißt, eine höhere Bitbreite führt zu einer feineren Abstufung der bei der A-D-Wandlung erhaltenen Werte.

Wenn man die Bitbreite bei der Quantisierung und die Sampling-Rate hoch genug wählt, entspricht das Ausgabesignal weitgehend dem Originalsignal. Der notwendige Datenfluß für ein akustisches Signal läßt sich aus dem Produkt der Sampling-Rate (Samples pro Sekunde) und der genutzten Bitbreite (Bits pro Sample) ermitteln. Da der Datenfluß durch die Datenübertragungsrates sowie durch die Speicherkapazität beim Speichern der digitalen Informationen begrenzt wird, ist es notwendig, die anfallenden Daten durch eine Codierung zu verringern, ohne jedoch die Qualität des Sprachsignals unter ein bestimmtes Niveau sinken zu lassen. Zur Codierung der Daten gibt es Ansätze (44), die als Grundlage lineare prädiktive Algorithmen (linear predictive coding- LPC) haben. Andere Codierungsansätze nutzen frequenzabhängige Algorithmen unter Berücksichtigung des Hörmodells, um die durch die Quantisierung hervorgerufenen Fehler unterhalb der Hörschwelle zu lassen (6, 10, 39).

Letztendlich führt eine Verringerung des Datenflusses zu einem komplexen Analyse-/Codierungsprozess und ab einer bestimmten Schwelle zu einer signifikanten Verschlechterung des Audiosignals. Die sogenannte „toll quality“ (vergleichbar mit einer guten Telefonverbindung über eine große Entfernung) kommt mit einer Bandbreite von 4000 Hz (mit

einer Samplingrate von 8000 Hz) und 8 Bit pro Sample bei direktem Sampling, d.h. unkomprimierter Datenübertragung auf einen Datenfluß von 64000 Bits/s (44). Zur Zeit kann durch Nutzung der Codierung bei Erhaltung der „toll quality“ die Datenübertragungsrate auf 8000 Bits/s gesenkt werden (11). Mit Qualitätsverlust des rekonstruierten Signals, aber noch verständlicher Sprache kann die Bit-Rate bis auf 2000 Bits /s gesenkt werden (11). Das gesampelte Sprachsignal enthält viele Daten, die für die eigentliche Information nicht immer notwendig sind. Bei einem vorgelesenen Text entsteht schätzungsweise eine Text-Datenrate von ca. 100 Bits/s (44). Eine höhere Effektivität bei geringerem Datenfluß wäre möglich, wenn Textdaten übertragen, zur Ausgabe in Parameter eines Sprachsynthesizers umgewandelt und über einen D-A-Wandler akustisch ausgegeben würden (Abb. 4).



**Abb. 4 : Sprachsynthesizer - Schema**

Zur Sprachsynthese ist ein digitales Lexikon des Ausspracheregelerkes mit allen Besonderheiten wie Aussprachegeschwindigkeit, Stimmhöhe, Lautstärke, etc. notwendig. Die Senkung des Datenflusses bei der Übertragung von Textdaten könnte nur durch erhöhten Rechenaufwand des empfangenden/auswertenden Systems erkaufte werden. Bei der Entwicklung eines sprachabhängigen Kommunikationssystems müssen also folgende Bedingungen berücksichtigt werden:

- der Aufwand der Analyse/Syntheseberechnungen
- der entstehende Datenfluß
- die notwendige Ausgabequalität
- die Flexibilität / Komplexität des Systems.

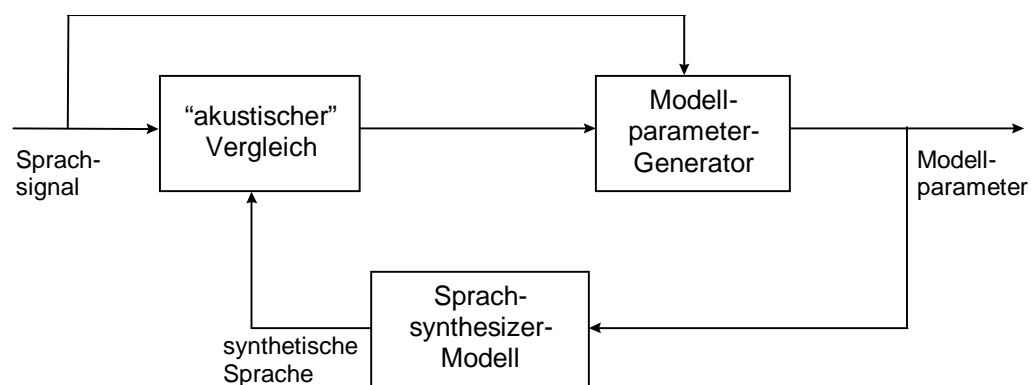
So wäre die Ausgabequalität von gesampelter Sprache bei hoher Datenübertragungsrate sicher optimal, es würde aber eine große zu verwaltende Datenmenge anfallen. Bei Nutzung der Sprachsynthese wiederum wäre ein erhöhter Rechenaufwand mit einem von der notwendigen Flexibilität abhängigen, umfassenden Regelwerk notwendig, wobei sich die Qualität der Sprachausgabe von fast unverständlicher „Roboter“-Sprache in den Anfängen der Sprachsynthese zu einer heutzutage gut verständlichen fast natürlich klingenden synthetischen Sprache entwickelt hat. Die Entwicklung der Sprachanalyse und -synthese steht in engem Zusammenhang mit den Gebieten der Akustik, der digitalen Signalverarbeitung, der Sprachforschung und der Psychologie. In der nächsten Zeit werden große Fortschritte vor allem in den folgenden Bereichen erwartet (44):

**Sprachmodell:** Darstellung sprachlicher Strukturen und phonetischer Elemente (Silben, Betonung,...) im akustischen Sprachsignal.

**Akustisches Modell:** Das lineare Modell der Spracherzeugung (Abb. 2, Seite 12) stellt nur eine einfache Basis für Sprachanalyse/-Codierung dar. Es läßt alle nichtlinearen Vorgänge außer acht. Aktuelle Forschungen (11, 26) auf diesem Gebiet beschäftigen sich mit der Entwicklung neuer nichtlinearer Modelle, die auf der Modulationstheorie, auf Fraktalen und der Chaostheorie basieren.

**Hörmodelle:** Forschungen auf dem Gebiet des Hörvorgangs sowie des Gehörs führen zu Modellen, die eine dramatische Verbesserung auf dem Gebiet der hochqualitativen Codierung (11) bewirken können.

**Analyse durch Synthese:** Das unveränderte Sprachsignal wird mit dem nach Parametermittlung synthetisch erzeugten Signal (bei vorgegebenen Ausschlußkriterien) verglichen, bei Ungleichheit der beiden Signale kommt es zu einer optimierten Parametererzeugung mit erneutem Vergleich (Abb. 5).



**Abb. 5 : Analyse durch Synthese**

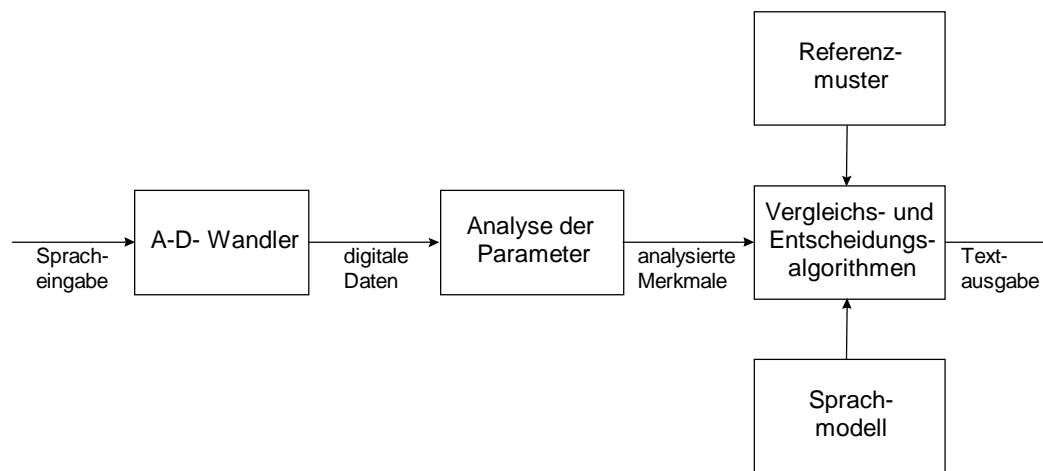
Dieses Prinzip stellt die Basis für verschiedene Codiersysteme dar (44). Eine Verbindung dieses Algorithmus mit genaueren Vergleichsmethoden, die z.B. auf dem Hörmodell und anderen Signalmodellen beruhen, wäre eine Möglichkeit der optimierten Codierung.

### 1.2.5 Spracherkennung und Sprachverständnis

Die Extraktion der Textinformation aus dem Sprachsignal ist eines der Hauptprobleme bei der Verständigung zwischen Mensch und Maschine. Es gibt viele grundsätzliche Gemeinsamkeiten zwischen der Spracherkennung und der Sprachanalyse/-synthese. Während bei der Sprachsynthese das Wissen über Spracherzeugung und -wahrnehmung sowie über die Ausdrucksform der sprachlichen Strukturen in der natürlichen Sprache letztendlich zur

Erzeugung eines Sprachsignals genutzt wird, dienen diese grundsätzlichen Kenntnisse bei der Spracherkennung zur Isolierung der eigentlichen Textinformation.

Abb. 6 (Seite 16) zeigt den grundsätzlichen Aufbau eines Spracherkennungssystems. Nach der Digitalisierung der analogen Audiodaten werden die digitalen Daten über ähnliche Algorithmen wie bei der Analyse/Synthese (z.B. Lineare prediktive Analyse, Filtertechniken,...) parameterisiert (40). Die so erhaltenen Merkmale erleichtern in den nachfolgenden Schritten die Decodierung des Audiosignals. Im Zentrum der Spracherkennung steht der Vergleich der Merkmale des gesprochenen Textes mit den während einer Trainingsperiode ermittelten Referenzdaten.



**Abb. 6 : Spracherkennungssystem - Aufbau**

Verschiedene Algorithmen (z.B. cepstrum distance measures, dynamic time warping (DTW), hidden Markov models (HMM)) werden zum Vergleich genutzt (40). Bei hochentwickelten Systemen werden noch Grammatik und Sprachmodelle in die Entscheidungsfindung einbezogen.

Die Spracherkennungssysteme werden nach ihren Fähigkeiten eingeteilt. Es gibt sprecherabhängige Systeme, bei denen Referenzdaten zum Vergleich herangezogen werden, die in einem Trainingsprozeß von jedem einzelnen Nutzer erstellt werden müssen. Diese Systeme sind in Ihrer Nutzeranzahl relativ begrenzt. Demgegenüber existieren sprecherunabhängige Spracherkennungssysteme, die eine unbegrenzte Nutzeranzahl ermöglichen. Einige Systeme erkennen eine große Anzahl von Wörtern oder Wortgruppen, andere nur einige Worte bzw. Zahlen. Oft ist nur die Einzelworterkennung möglich, das heißt hinter jedem gesprochenem Wort muß eine definierte Pause gemacht werden, damit einzelne Worte voneinander abgegrenzt werden können. Weitaus komplizierter ist die kontinuierliche Spracherkennung, wobei im Redefluß Pausen nicht zwingend notwendig sind. Ein System mit geringem Wortschatz, Einzelworterkennung und Benutzerabhängigkeit wäre relativ einfach entwickelbar, während ein System mit kontinuierlicher Spracherkennung und großem Voka-

bular (bei gleichzeitiger Nutzerunabhängigkeit) der natürlichen Sprechweise am ehesten gerecht wird.

Natürlich wird die Komplexität eines Systems der zu erfüllenden Aufgabe angepaßt werden müssen. Während zum Setzen einzelner Schalter ein einfach strukturiertes System genügt, ist für ein Freitextdiktiersystem ein weitaus größerer Aufwand notwendig.

Bei der weiteren Entwicklung von Spracherkennungssystemen wird insbesondere das Sprachmodell eine große Rolle spielen. Ähnlich wie bei der Sprachsynthese ist das Verständnis für die Darstellung der Sprachstrukturen und -bestandteile im akustischen Sprachsignal für die Verbesserung der Spracherkennung wichtig. Insbesondere die Algorithmen für die Parameteranalyse und die Vergleichsoperationen würden davon profitieren. Die stetige Weiterentwicklung der Mikroprozessortechnik wiederum ermöglicht durch steigende Rechenleistung erst den Einsatz neu entwickelter komplizierter Algorithmen in einem angemessenem Zeitrahmen.

#### 1.2.6 Stand der Entwicklung und Forschung in der digitalen Sprachverarbeitung

Als größte kommerzielle Anwendung wurde die digitale Spracherkennung im Operator-Service verschiedener Telefongesellschaften (zuerst in den USA und in Kanada) eingesetzt (35), anfangs, um die Antworten „Yes“ und „No“ zu unterscheiden. In der weiteren Entwicklung wurde das Vokabular erweitert. Durch die Automatisierung sparten die Telefongesellschaften jedes Jahr Hunderttausende Dollar ein (45).

Als Hilfe für Behinderte gibt es Geräte (Telefone, Betten, Rollstühle,...), die unter Nutzung einer digitalen Spracherkennung mit geringem Vokabular steuerbar sind (45). Hörhilfen, die mit einer Spracherkennung arbeiten, sind zur Zeit noch in der Experimentalphase (24).

Seit Anfang der 90-er Jahre gibt es verschiedene Applikationen zur Bedienung von Computern (35). Seit Mitte der 90-er Jahre werden von verschiedenen Herstellern digitale Spracherkennungssysteme mit unterschiedlichem Vokabelumfang angeboten (15, 20, 29, 31, 35).

Die Ergebnisse der Forschung auf dem Gebiet der Spracherkennung werden auch militärisch genutzt. Um beide Hände freizuhaben, wird die Spracherkennung hauptsächlich zur Steuerung von Geräten eingesetzt. Sie wird auch zur Datenerfassung genutzt (50).

Die nächste Weiterentwicklung auf dem Gebiet der digitalen Sprachverarbeitung wird in der Verbesserung und Erweiterung der bisherigen Anwendungen bestehen (35).

Von der digitalen Spracherkennung im Zusammenhang mit der Spracherzeugung und der Stimmcodierung werden weitreichende Entwicklungen erwartet (1, 12, 23). So wurde schon 1983 ein Experiment zu einem multilingualen Telefonsystem vorgestellt (11). Wilpon (51)

formuliert die futuristisch anmutende Vision einer natürlichen, kontinuierlichen Kommunikation zwischen Menschen und Maschinen in irgendeiner Sprache, so daß der Zugriff auf Informationen oder Dienste bzw. der Informationsaustausch in nahezu jeder Sprache verfügbar sei. Schon auf der TELECOM'91 wurde der INTERTALKER präsentiert, ein automatisches Übersetzungssystem, daß ein integriertes sprecherunabhängiges Spracherkennungssystem für Englisch und Japanisch enthielt und den erkannten Text über einen Sprachsynthesizer in Englisch, Japanisch, Französisch und Spanisch wiedergeben konnte (19).

### **1.3 Einsatz der digitalen Spracherkennung in der Medizin**

Vereinzelte Veröffentlichungen zeigen erste Versuche des Einsatzes der digitalen Spracherkennung im Bereich der Medizin. So berichtete Shilitoe (46) über einen DNA-Sequenzeditor, der außer der reinen Sequenzeingabe über die Tastatur auch die Möglichkeit bot, eingegebene Sequenzen anzuhören und Sequenzen über Spracheingabe zu editieren. Das System beschränkte sich auf einen sehr geringen Wortschatz („adenine“, „cytosine“, „guanine“ und „thymidine“) und erforderte eine nutzerspezifische Trainingsphase, in der diese Worte trainiert wurden. Während des Diktates der Sequenz bestätigte der Sequenzeditor (SE) jedes erkannte Wort mit dem zugehörigen Buchstaben (A, C, G, und T) und trug den Buchstaben in die Sequenz ein. So war die Kontrolle der Eingabe gewährleistet, gleichzeitig konnten Hände und Augen des Untersuchers der Sequenz auf dem Sequenz-Gel oder dem Papier folgen. Kolles (20) berichtete über den Einsatz des DragonDictate<sup>TM</sup>-30K-Systems (DD) im Einsatz in der diagnostischen Pathologie. Bei diesem System handelte es sich um ein sprecherabhängiges Einzelworterkennungssystem. Bei der getesteten Version handelte es sich um ein DOS-Programm, das die gesprochenen Worte in Tastatureingaben umwandelt. Das Programm adaptierte sich während der Nutzung an die Sprachgewohnheiten des Einzelnen, so daß je nach vorhandenem Vokabular über einen längeren Zeitraum schlechte Erkennungsraten vorlagen und ein erhöhter Korrekturaufwand notwendig war. Als besondere Stärke in der täglichen Routine wurde die Möglichkeit der Nutzung von Textbausteinen hervorgehoben. Bei schwierigeren Diktaten mit selten genutztem Vokabular wurde die Nutzung des DD jedoch nicht empfohlen.

Schon 1988 berichtete Hansen (15) über erste Versuche des Einsatzes eines Spracherkennungssystems (VoiceRad) in der radiologischen Diagnostik. Er hob hervor, daß dieses System durch seine Sprecherabhängigkeit und den Aufwand der Kontrolle des erkannten Textes für den Routinebetrieb nicht einsetzbar wäre. Es stellte sich heraus, daß die Zeit, die

zur Erstellung eines Befundtextes mittels des VoiceRad-Systems benötigt wurde, proportional zu der Komplexität und Abnormität der befundeten Bilder anstieg.

Der 1993 in der radiologischen Diagnostik getestete Prototyp des Spracherkennungssystems Tangora (31) stellte ebenfalls ein sprecherabhängiges Einzelworterkennungssystem dar. Als großer Vorteil wurde das sofortige Vorliegen des schriftlichen Befundes gewertet. Demgegenüber stand der erhöhte Arbeitsaufwand durch Einzelwortsprechweise mit verringerter Diktiergeschwindigkeit sowie die notwendigen Korrekturen. Die durchschnittliche Erkennungsrate bei 68 erfaßten Befunddiktionen betrug 93%. Trotz des erhöhten Arbeitsaufwandes im Vergleich zur herkömmlichen Diktatpraxis wurde der Vorteil des schnellen Vorliegens des schriftlichen Befundes in den Vordergrund gestellt.

1994 berichtete Teplitz (49) über die Implementation des Spracherkennungssystems VoicePATH (Version 4.x) innerhalb eines Systems zur Erstellung intraoperativer pathologischer Befunde (Automatic Speech-Recognition Anatomic Pathology Reporting [ASAP] ). Das VoicePATH verfügte über einen Wortschatz von 5000 Worten. Nach Fertigstellung eines Befundes konnte dieser automatisch an den Überweiser gefaxt werden. Desweiteren beinhaltete das komplexe System einen Sprachsynthesizer, über den schriftliche Befunde in Sprache umgesetzt wurden. So konnten die Kliniker den Befund über normale Telephone bzw. über Mobiltelephone abhören.

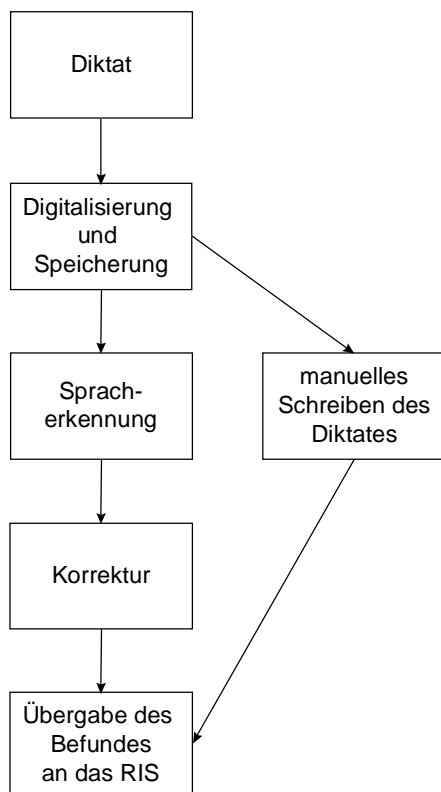
## 2 Zielstellung

Die computergestützte digitale Spracherkennung bietet ein großes Potential zur beschleunigten Erstellung detaillierter schriftlicher Befunde. Zur Evaluierung der zur Zeit vorhandenen Möglichkeiten wurde das digitale Spracherkennungssystem SP 6000 im Einsatz in der täglichen Routinearbeit der radiologischen Diagnostik getestet. Dabei sollte zum einen die Erkennungsrate und deren Dynamik im Einsatz unter normalen Arbeitsbedingungen über einen längeren Zeitraum untersucht werden. Desweiteren sollte die direkte Einbindung des Systems in das bestehende RIS sowie der Arbeitsaufwand bei der Einarbeitung und Nutzung des Systems betrachtet werden.

Um die Vor- und Nachteile der digitalen Spracherkennung beurteilen zu können, erfolgte ein Vergleich des Spracherkennungssystems mit der momentan üblichen Befundungsweise.

## 3 Material und Methoden

### 3.1 Das Spracherkennungssystem SP 6000



**Abb. 7 : Befundung mit digitaler Spracherkennung - Arbeitsablauf**

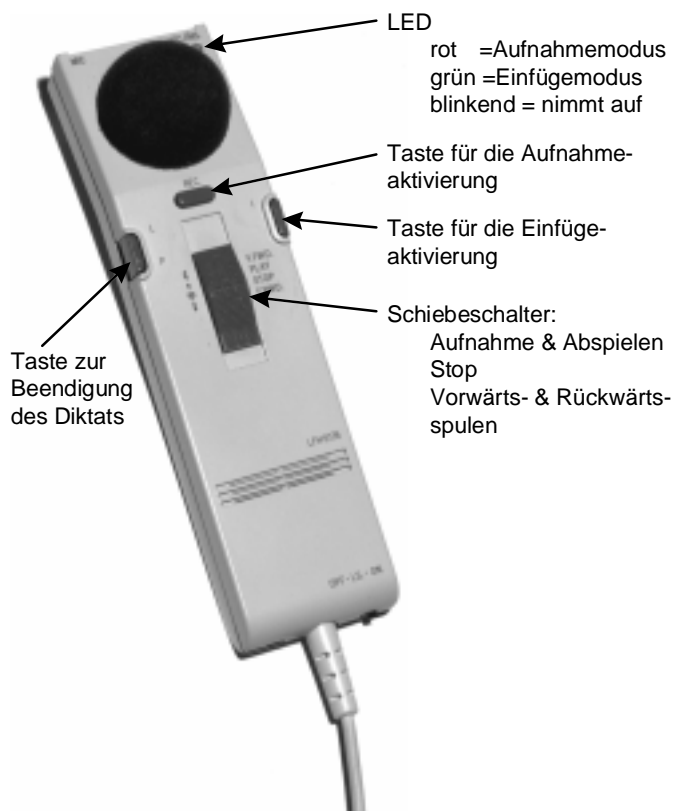
Das getestete SP 6000 ist ein von Philips entwickeltes Spracherkennungssystem. Bei diesem System wird über eine Kopplung von Software und speziell entwickelter Hardware mittels kontextspezifischer kontinuierlicher Spracherkennung gesprochener Text in geschriebenen Text umgewandelt. Dabei ist es möglich, über spezielle Steuerworte während des Diktierens eine Trennung von Daten ( die bei der Übergabe an manche RIS-Systeme in spezielle Tabellenfelder eingetragen werden bzw. auch nur zur Identifizierung des Diktates während der Korrektur dienen) vom eigentlichen Befund sowie eine Formatierung des Textes zu erreichen.

Abb. 7 verdeutlicht den Arbeitsablauf bei der Erstellung eines Befundes. Das Diktat wird über

ein standardisiertes Philips-Diktatmikrofon aufgenommen, über eine philipseigene Digitalisierungskarte in ein digitales Sprachsignal umgewandelt und auf einer Festplatte abgespeichert (siehe 3.1.1; Seite 21). Das digital gespeicherte Sprachsignal kann nun der digitalen Spracherkennung zugeführt oder auch abgehört und manuell geschrieben werden. Im automatisierten Ablauf wird das gespeicherte Diktat mit Hilfe einer philipseigenen Acceleratorkarte in geschriebenen Text umgewandelt (siehe 3.1.2; Seite 22). Nach der Umwandlung des gesamten Diktates steht der geschriebene Text zur Korrektur bereit. Nach der Korrektur (siehe 3.1.3; Seite 22) wird das Diktat an das Radiologische Informationssystem übergeben und dort an die Befundschreibung übergeben (siehe 3.1.4; Seite 23).

### 3.1.1 Diktat

Das Diktieren erfolgt über ein an die Digitalisierungskarte angepaßtes Mikrofon (Abb. 8), das sich im Aussehen und in der Handhabung nicht von einem Standarddiktatmikrofon unterscheidet. Beim Starten des Diktiervorganges mittels der Record-Taste wird auf dem



**Abb. 8 : Diktatmikrofon des SP 6000**

öffnet sich ein neues Auswahlfenster. Hier kann eine Priorität des Diktats sowie eine genaue Bezeichnung des Diktats (wenn notwendig) vergeben werden. Nach Bestätigung

Bildschirm ein Fenster geöffnet, das die aktuelle Diktatlänge, die aktuelle Position im Diktat sowie den momentanen Status (z.B. Aufnahme, Pause, Wiedergabe, Einfügen) anzeigt. Desweiteren bietet sich in diesem Fenster die Möglichkeit, das Diktat zu beenden, zwischenzuspeichern (um später weiter zu diktieren) oder das Diktat abubrechen. Es besteht keine Notwendigkeit, während des Diktierens Pausen zwischen den einzelnen Worten einzufügen, da das System für die kontinuierliche Spracherkennung entwickelt wurde. Bei Beendigung des Diktats

der Einstellungen bekommt der diktierter Befund den Status „Frei Erkennung“ und steht somit dem eigentlichen Spracherkennungsmodul zur Verfügung.

Beim Befunden wurden generell die mitgeteilten klinischen Daten, der eigentliche Befund und das Untersuchungsergebnis diktiert, so daß auch klinische Begriffe außerhalb des direkt radiologischen Kontextes Verwendung fanden.

#### 3.1.2 Spracherkennung

Sobald ein Diktat den Status „Frei Erkennung“ erhalten hat, wird es vom Spracherkennungsserver (einem Teilmodul des Systems) nach Priorität in die Liste der zu erkennenden Diktate eingefügt und bearbeitet. Dabei wird das Diktat mittels der phonetischen Referenz des Diktierers in einzelne Worte zerlegt. Der phonetisch erkannte Begriff wird mit den vorhandenen Worten im Lexikon verglichen. Wenn einem Wort mit hoher Wahrscheinlichkeit ein Eintrag im Lexikon zugeordnet werden kann, wird anschließend der Bezug zu den vorher erkannten Worten geprüft. Ergibt sich bei dieser kontextspezifischen Überprüfung eine hohe Unwahrscheinlichkeit für das erkannte Wort, so erfolgt eine erneute Suche im Lexikon nach einem Wort mit ähnlicher Phonetik. Das Wort mit der höchsten phonetischen und kontextspezifischen Wahrscheinlichkeit wird als erkannt in den geschriebenen Text übernommen. So sind z.B. die Worte „Haus“ und „Maus“ phonetisch sehr ähnlich. Bei unsauberer Sprechweise kann das System nur aus dem Kontext ermitteln, welches Wort das richtige ist. Das bedeutet, der Begriff „das Haus“ ist wahrscheinlicher als „das Maus“, so daß auch bei phonetisch ähnlich klingenden Begriffen durch ihre Kontextspezifität eine gute Unterscheidbarkeit vorliegt.

Wenn der letzte Terminus des Diktats erkannt wurde, erhält das Diktat den Status „Frei Korrektur“.

#### 3.1.3 Korrektur

Da bei der Erkennung Fehler auftreten können, muß das Befunddiktat nach der kompletten Erkennung korrigiert werden. Zu diesem Zweck wird der Korrektureditor mit dem entsprechenden Diktat geöffnet. Eventuell vorhandene falsch erkannte Worte (einschließlich Zahlen und Steuerworte) müssen nun korrigiert werden. Der Editor ermöglicht das Abhören des diktierten Befundes mit gleichzeitiger Markierung des zugeordneten Textes, so daß eine direkte Kontrolle möglich ist. Korrigierte Worte werden mit einer anderen Textfarbe (grün) kenntlich gemacht.

Sachlich falsch diktierte Worte oder Befundteile dürfen nicht im Korrektureditor verändert werden. Würde man falsch diktierten, aber richtig erkannten Text an dieser Stelle verändern, so würde das System beim nächsten akustischen Training Sprachsignal mit dazu unpassendem Text verarbeiten, was eine deutliche Verschlechterung der phonetischen Referenz und damit der Erkennungsrate zur Folge hätte.

Nach Beendigung der Korrektur kann das Diktat direkt an das jeweilige Befundungssystem übergeben werden. Bei einer späteren (automatisierten) Verarbeitung des korrigierten Textes werden die korrigierten Worte vom Spracherkennungsserver mit den vorhandenen Einträgen im Lexikon verglichen. Ist ein Wort nicht im Lexikon vorhanden, so wird es in die Liste der neuen Worte aufgenommen.

#### 3.1.4 Übergabe des Befundtextes

Nach Beendigung der Textkontrolle wird der Text im RTF-Format gespeichert und an das hausinterne Befundungssystem übergeben. Wir nutzten das Radiologische Informationssystem MEDORA. Die Befundschreibung wird bei diesem System über Word für Windows 6.0 realisiert. Zur Übernahme des Befundtextes aus der Spracherkennung ist es notwendig, die entsprechende Untersuchung des jeweiligen Patienten aufzurufen. Der Text wird an der aktuellen Cursorposition formatiert, d.h. mit Umsetzung der diktierten Steuerworte (z.B. Neue\_Zeile, Fett\_Anfang, Fett\_Ende,...), eingetragen.

Erst jetzt, nach Übergabe des diktierten Befundes, ist eine inhaltliche Veränderung des erkannten Textes (Zusätze, Streichungen, etc.) zulässig.

#### 3.1.5 Lexikon

Das Lexikon und auch das kontextspezifische Sprachmodell ist speziell an den radiologischen Kontext angepaßt. Das Wörterbuch enthält ca. 55000 Einträge. Durch die Nutzer können weitere Wörter bis zu einer Gesamtanzahl von 65000 Worten eingetragen werden. Nach Abschluß der Korrektur wird das korrigierte Diktat automatisch auf unbekannte, d.h. im Lexikon noch nicht vorhandene Worte, überprüft. Mittels des Moduls Vokabelmanager (VOM) können diese neuen Worte überprüft und gegebenenfalls in das Lexikon übernommen werden. Bei der Überprüfung der neuen Worte besteht auch die Möglichkeit der Sonderbehandlung. Sonderbehandlung bedeutet hier, daß Worte in ihrer Phonetik nicht mit denen der deutschen Sprache übereinstimmen (z.B. Anglismen) oder ein Wort bzw. eine Wortgruppe durch eine phonetisch nicht kongruente Zeichenfolge (z.B. gesprochen: "römisch\_Eins", geschrieben „I“) korrekt in das Lexikon eingebracht werden können.

Werden nach Erreichen der maximalen Wortanzahl weitere Einträge gemacht, so werden die Worte mit der geringsten Nutzungshäufigkeit entfernt, um für die neuen Einträge Platz zu schaffen.

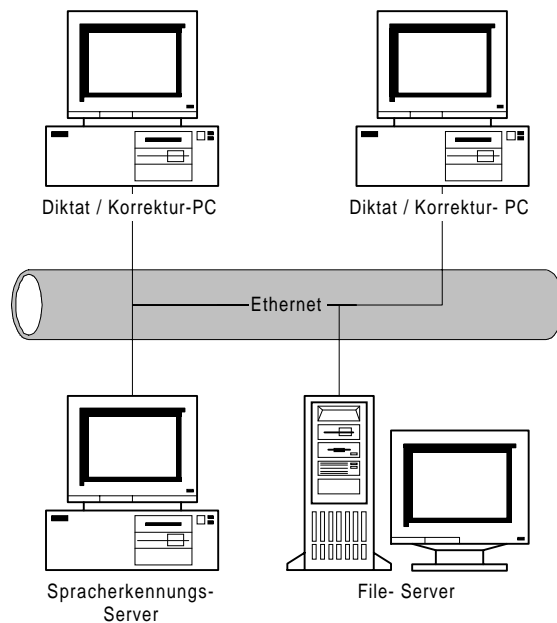
#### 3.1.6 Initialtraining und Adaptation

Das SP 6000 ist ein benutzerspezifisches Spracherkennungssystem. Das heißt, für jeden Nutzer des Systems ist eine Anpassung des Systems notwendig. Die erste Anpassung, auch Initialtraining genannt, muß vor der ersten Nutzung der Spracherkennung erfolgen. Zu diesem Zweck existieren 25 vom Hersteller vorgegebene Befundtexte. Diese Texte enthalten einen Großteil der in der Radiologie üblichen Worte und Wortgruppen und müssen wörtlich mit Formatierung und Notation von dem neuen Nutzer dem System vorgelesen werden. Der Systembetreuer sollte während des Lesens der Diktate anwesend sein, um Lesefehler sofort korrigieren zu lassen. Nachdem alle Diktate gelesen und kontrolliert wurden, kann das Initialtraining gestartet werden. Dabei erstellt das System aus dem ihm bekannten Text der Diktate und aus dem vorliegendem Sprachsignal ein benutzerspezifisches Sprachreferenzfile. In diesem Referenzfile ist die persönliche Aussprache der einzelnen Phoneme der deutschen Sprache abgelegt. Zusätzlich wird ein persönliches Sprachmodell für den Kontext der Radiologie angelegt. Dieses persönliche Sprachmodell enthält Informationen über die Nutzung (Wahrscheinlichkeit) von Wortfolgen im Sprachgebrauch des einzelnen Nutzers.

Da die Aussprache von gelesenem Text nicht vollständig der Aussprache des freien Diktats entspricht, sind weitere Trainingszyklen notwendig. Diese Trainingszyklen, auch Adaptation genannt, können durchgeführt werden, sobald eine ausreichende Menge Sprachsignal (d.h. Befundungsdiktate) vorliegt. Eine Adaptation sollte erst durchgeführt werden, wenn mindestens ca. 2 h Sprachsignal vorhanden sind. Die Adaptation unterteilt sich in ein akustisches Training sowie eine Vokabelaktualisierung. Bei dem akustischem Training wird aus den vorhandenen Befunden (Sprachsignal und korrigierter Text) ein neues Sprachreferenzfile erstellt. Das akustische Training ist wiederholt notwendig, da sich die Phonetik des Einzelnen im Laufe der Zeit ändern kann (z.B. verändern Zahnextraktionen, -ersatz den akustischen Resonanzraum).

Die Vokabelaktualisierung (Vocabulary Update) beinhaltet eine Aktualisierung des Lexikons sowie des Sprachmodells, wodurch eine bessere Anpassung an den persönlichen Sprachgebrauch erfolgt.

### 3.1.7 Konfiguration des Systems



Die Installation des SP 6000 wurde in das im Hause vorliegende Netzwerk eingefügt. Die Hardwarekonfiguration (Abb. 9) besteht aus

- 1 Fileserver
  - 1 Server zur Spracherkennung
  - 2 Diktat- und Korrektur-PC's
- die über eine Netzwerkverbindung mittels Novel Netware verbunden sind.

Der Fileserver dient zur zentralen Speicherung aller Daten des Spracherkennungssystems, wie Programmdaten, Diktate (digitales Sprachsignal sowie erkannte und korrigierte Texte), benutzerspezifische Daten (Sprachreferenzfile, Sprachmodell) und des Lexikons.

#### Abb. 9 : Hardwarekonfiguration

Der Spracherkennungsserver ist mit einer speziellen Spracherkennungshardware (Acceleratorkarte) ausgestattet. Mit Hilfe dieser Karte werden die fertigen, auf dem Fileserver gespeicherten Diktate verarbeitet und die erkannten Texte auf dem Fileserver abgespeichert. Die Diktatplätze sind mit einer speziellen Digitalisierungskarte ausgerüstet, die außer der Digitalisierung des Sprachsignals auch die Wiedergabe des diktierten Befundes sowie die Verarbeitung der Steuersignale des Mikrofons und des Fußschalters ermöglicht. Während sich die Diktat-/Korrektur-PC's direkt im Arbeitskern befinden, stehen der Fileserver und der Spracherkennungsserver in einem zentralen Computerraum. Als Betriebssystem wird auf den Diktat-PC's und auf dem Spracherkennungsserver MS-DOS 6.22 und Windows for Workgroups 3.11 genutzt. Auf dem zentralen Fileserver läuft Novel Netware 3.12 (Lizenz für 5 Benutzer).

## 3.2 Testpersonen

Arzt 1

- 4 Jahre Berufserfahrung
- geringe Diktiererfahrung mit herkömmlichen Systemen
- Tastaturbenutzung: hohe Schreibgeschwindigkeit
- sehr gute Computerkenntnisse

Arzt 2

- 4 Jahre Berufserfahrung
- 4 Jahre Diktiererfahrung mit herkömmlichen Systemen
- Tastaturbenutzung: geringe Schreibgeschwindigkeit
- gute Computerkenntnisse

Arzt 3

- 0,5 Jahre Berufserfahrung
- keine Diktiererfahrung
- Tastaturbenutzung: mäßige Schreibgeschwindigkeit
- sehr gute Computerkenntnisse

### 3.3 Erfasste Werte

Zur Beurteilung der Funktionalität des Spracherkennungssystems wurde die Fehlerrate nach dem Erkennungsvorgang bei jedem Diktat ermittelt. Außerdem wurden das Datum des Diktats, die Diktatlänge sowie der Zeitaufwand der Kontrolle/Korrektur in Sekunden erfaßt. Zusätzlich wurde die Art der dem Befundungsdiktat zugrundeliegenden Untersuchung notiert. Die Diktat-PC's standen zum Zeitpunkt der Testung im Arbeitskern. Der Arbeitskern war ein zentraler Befundungsraum, in dem sich zum Zeitpunkt der Datenerfassung 6 Befundungsplätze, 5 RIS-Arbeitsplätze, 3 Entwicklungsautomaten und ein Belichtungsautomat befanden. Zwei der Befundungsplätze konnten von 2 Befundern gleichzeitig genutzt werden. In direktem Anschluß war der Arbeitskern von 6 verschiedenen Untersuchungsräumen sowie einer Dunkelkammer umgeben, so daß während der Befundung gleichzeitig der direkte Kontakt zur Untersuchung bestand.

Das heißt, die Werte wurden nicht unter schallisolierten Laborbedingungen ermittelt, sondern im Routinebetrieb unter Einwirkung alltäglicher Störgeräusche (Telefonklingeln, Arbeitsgeräusche der umstehenden Entwicklungs- und Belichtungsautomaten, Türenschlagen, Gespräche,...).

Als Vergleichsmöglichkeit wurden durch zwei der Testpersonen (Arzt 1 und Arzt 3) zusätzlich Daten von konventionell geschriebenen Befunden erhoben. Dabei wurde die Zeitdauer des Schreibens, die Art der Untersuchung und die eventuelle Nutzung von Textbausteinen erfaßt.

- Ermittlung der Fehlerrate

Zur Ermittlung der Fehlerrate jedes einzelnen Diktats wurde ein im Korrektureditor integriertes Werkzeug genutzt. Dieses berechnet aus der Anzahl der korrigierten

Worte im Verhältnis zu der Summe aller Worte des jeweiligen Diktats eine prozentuale Fehlerrate.

Die Fehlerrate wurde jeweils nach Beendigung der Kontrolle/Korrektur bestimmt.

- Ermittlung der Diktatlänge und der Korrekturzeit

Die Diktatlänge wurde vom SP 6000 während des Diktats aktualisiert und bei jedem Zugriff (Diktieren/Korrektur) angezeigt. Dabei wurden längere Sprechpausen (in Abhängigkeit von der eingestellten Empfindlichkeit des Mikrophons) automatisch nicht berücksichtigt, da bei Sprechpausen eine automatische Unterbrechung der Aufzeichnung erfolgte.

Die Korrekturzeit wurde definiert als Zeit zwischen dem Aufruf des Korrektureditors und der Beendigung der Korrektur durch Übergabe des Textes an das RIS bzw. durch Start der Fehlerratermittlung. Die Zeit zur Kontrolle und gegebenenfalls zur Korrektur des Diktats wurden mittels Stoppuhr bestimmt.

- Ermittlung der Schreibdauer

Die Schreibdauer wurde mittels Stoppuhr erfaßt. Die Stoppuhr wurde zu Schreibbeginn gestartet, das heißt, nachdem der Patient im RIS aufgerufen worden war. Die Zeit wurde bei Beendigung des Schreibens, also zum Zeitpunkt des Druckens b.z.w. Speicherns gestoppt.

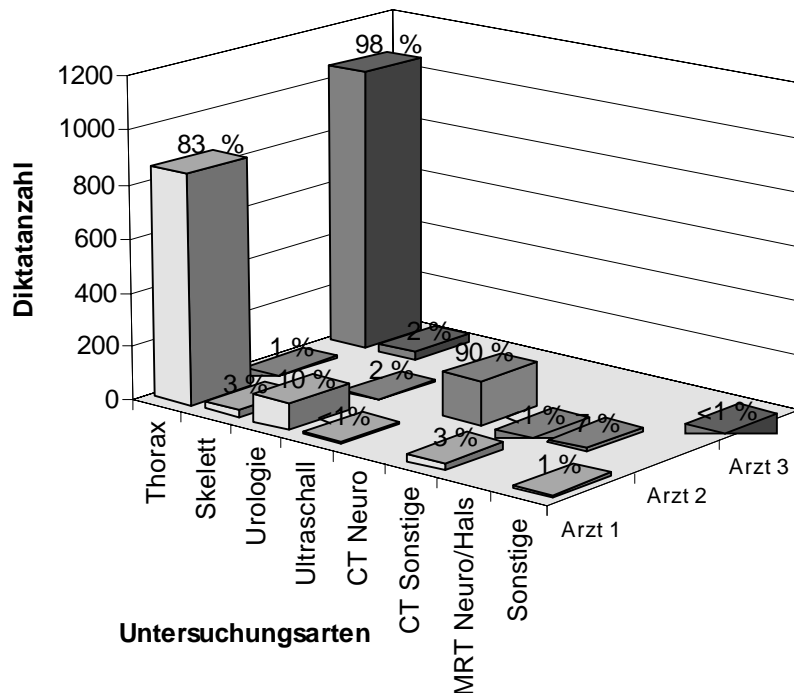
### 3.4 Datenmenge

Gesamt	Thorax	Skelett	Thorax + Skelett	Magen/ Darm	Urologie	Urologie + Zweitunters.	HNO
<b>2305</b> (100 %)	1921 (83 %)	53 (2 %)	8 (> 0.5 %)	3 (> 0.5 %)	104 (5 %)	2 (> 0.5 %)	1 (> 0.5 %)
Ultraschall	CT Thorax	CT Abdomen	CT Thorax + Abdomen	CT Neuro	CT Angio/ Hals/ Kopf	MRT Hals	MRT Neuro
8 (> 0.5 %)	5 (> 0.5 %)	11 (> 0.5 %)	9 (> 0.5 %)	164 (7 %)	4 (> 0.5 %)	2 (> 0.5 %)	10 (> 0.5 %)

**Tabelle 1 : Verteilung der erfaßten Befundungsdiktate**

Tabelle 1 zeigt die Aufschlüsselung der erfaßten Diktate auf die zugrundeliegenden Untersuchungen. Es wurden Daten zu 2305 Diktaten erfaßt. Die große Anzahl berücksichtigter Thoraxuntersuchungen (83 %) ist durch die willkürliche Auswahl dieses Arbeitsplatzes zur Testung des Spracherkennungssystems zu erklären. Durch die hohen Untersuchungszah-

len am Thoraxarbeitsplatz war ein zur Systemtestung und Datenerfassung ausreichend großes Diktataufkommen pro Tag und Arzt gesichert. Die Datenerfassung zu Befunden anderer Untersuchungen wurden durch den hausinternen Routine- und Rotationsbetrieb ermöglicht. So konnten z.B. durch Arzt 1 im Bereitschaftsdienst gelegentlich Daten zur CT- und Sonographiebefundung erhoben werden.



**Diagramm 1 : Verteilung der erfaßten Befundungsdiktate**

Diagramm 1 stellt die Aufteilung der Untersuchungen auf die einzelnen Testpersonen dar. Individuell wurden unterschiedliche Befunde erstellt. Bei Arzt 1 und Arzt 3 fällt eine Häufung der Thoraxbefundung auf (83 % bzw 98 % aller diktierten Untersuchungen), bei Arzt 2 tritt die Neuro - CT (90 %) in den Vordergrund. Während Arzt 1 und Arzt 3 seit Beginn der geplanten Testphase an der Datenerfassung beteiligt waren, arbeitete Arzt 2, der erst später an der Testung des Spracherkennungssystems teilnahm, an einem anderen Arbeitsplatz, so daß interindividuelle Unterschiede des Untersuchungsspektrums durch den internen Arbeitsablauf zu erklären sind. Durch die Nutzung der digitalen Spracherkennung im Bereitschaftsdienst und nach ausbildungsstandabhängigem Untersuchungsplatzwechsel wurden in geringer Anzahl Daten für verschiedene Untersuchungstechniken erfaßt.

Um eine Vergleichsmöglichkeit zur vorwiegend im Hause genutzten Befundungsart zu haben (der befundende Arzt schreibt unter Nutzung von Word für Windows den Befund selbst), wurden Daten von insgesamt 625 geschriebenen Befunden erfaßt (Arzt 1 und Arzt 3). Die Verteilung der erfaßten geschriebenen Befunde (Tabelle 2, Seite 29) wurde

durch den hausinternen Arbeitsablauf und durch den individuellen Ausbildungsstand bestimmt. Während von Arzt 3 Daten zu 142 (54 %) der schriftlichen Thoraxbefundungen erfaßt wurden, konnte Arzt 1 die Befundungsdaten eines weitaus größeren Untersuchungsspektrum liefern.

Gesamt	Thorax	Thorax + Zweitunters.	Skelett	Trauma	Abdomen Magen/ Darm	HNO	CT
<b>625</b> (100 %)	263 (42 %)	3 (<1 %)	9 (1 %)	181 (29 %)	5 (<1 %)	5 (<1 %)	2 (<1 %)
US Thorax	US Abdomen	US Abdomen + Zweitunters.	US NTX (+Zweitunters.)	US Niere	US FKDS Bein/ Hals	US Hals	US Weichteile/ Mamma
5 (<1 %)	106 (17 %)	23 (4 %)	2 (<1 %)	5 (<1 %)	8 (1 %)	5 (<1 %)	3 (<1 %)

**Tabelle 2 : Verteilung der erfaßten geschriebenen Befunde**

### 3.5 Statistische Auswertung

Die statistische Auswertung erfolgte unter Nutzung des Programme SPSS für Windows 6.0.1 und Excel 7.0.

#### 3.5.1 Mittelwert

Als Mittelwert wurde das arithmetische Mittel der jeweiligen Werte bestimmt.

#### 3.5.2 Standardabweichung

Die Standardabweichung wurde nach folgender Formel errechnet:

$$\text{STABW} = \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}}$$

#### 3.5.3 Ungepaarter Wilcoxon-Test (Mann-Whitney U-Test)

Da die Verteilung der Fehlerraten eindeutig nichtnormal war, wurde zur Beurteilung der statistischen Signifikanz einer beobachteten Veränderung der individuellen sowie interindividuellen Diktatfehlerrate nach jeder Adaptation der ungepaarte Wilcoxon- Rangsummen-Test gewählt. Bei dieser nichtparametrischen Prozedur werden zwei nichtverbundene Stichproben zum Test der Nullhypothese, daß die Rangsummen einer Variablen gleich sind, untersucht. Es sind keine Annahmen über die Form der Verteilung notwendig. Die absoluten Werte der Differenzen werden berechnet und vom kleinsten zum größten in eine Rangordnung gebracht. Die Teststatistik basiert auf den Rangsummen für die negativen und positiven Differenzen.

Zur Signifikanzentscheidung erhält man den Wert  $p$ , der das beobachtete Signifikanzniveau (observed significance level) ausdrückt.

Dieser gibt die Wahrscheinlichkeit an, daß die beobachtete Differenz der Mittelwerte zweier Stichproben nur zufällig zustande gekommen ist (statistischer Fehler 1. Art). Es wird üblicherweise festgelegt, daß der höchste akzeptable Wert für diese Wahrscheinlichkeit  $\alpha = 5\%$  ist, so daß bei einem Ergebnis von  $p < 0,05$  die Nullhypothese abgelehnt werden kann, da mit 95 % iger Wahrscheinlichkeit ein signifikanter Unterschied zwischen den Stichproben besteht.

## 4 Ergebnisse

### 4.1 *Arbeitsaufwand zur Installation, Netzwerkeinbindung und Kopplung des Systems an das bestehende RIS*

Der Zeitaufwand für die Einbettung des SP 6000 in obengenannter Konfiguration (3.1.7; Seite 25) wurde grundsätzlich auf 2 Tage veranschlagt. Das Spracherkennungssystem sowie die Novel Netware-Anbindung waren auf den gelieferten Rechnern vorinstalliert. Da die hausinterne Befundung mittels RIS (MEDORA)- gekoppeltem Winword 6.0 erfolgte und die Netzanbindung über das TCP/IP-Protokoll realisiert war, erhielten die Rechner eine gültige IP-Adresse. Nach der Installation der RIS-Datenbankverbindung erfolgte ein Test beider Systeme. Die Anbindung des Spracherkennungssystems an das RIS erfolgte zum Zeitpunkt der Installation über ein Winword-Makro, das den erkannten und korrigierten Text unter Abarbeitung der Steuerworte in das geöffnete Winword-Dokument einträgt. Insgesamt wurde für die Installation, inclusive Testung ca. 10 h benötigt.

### 4.2 *Zeitaufwand für das Initialtraining*

Der Zeitaufwand für das Lesen der Trainingsdiktate betrug für den neuen Nutzer und für den Systembetreuer in Abhängigkeit von der Lesegeschwindigkeit und der auftretenden Lesefehler des neuen Nutzers ca. 1,5 - 2,5 h. Das Initialtraining konnte nur auf dem Computer gestartet werden, in dem die Acceleratorkarte installiert war, in unserem Fall also auf dem Spracherkennungsserver. Der Rechner benötigte ca. 2 h für das Initialtraining. Zusätzlich zu dem eigentlichen Training fielen noch administrative Aufgaben an (Neueintrag und Paßwortvergabe, Nutzereinweisung) die mit ca 1 h zu Buche schlugen (Tabelle 3).

Lesen & Korrektur der Diktate	1,5 - 2,5 h	Initialtraining	2 h
Administration	1 h		
Personeller Aufwand : (2 Personen)	<b>2,5 - 3,5 h</b>	Maschineller Aufwand :	<b>2 h</b>

**Tabelle 3 : Zeitaufwand für das Initialtraining**

### 4.3 Zeitaufwand für die Betreuung des laufenden Systems

Bei der Administration des laufenden Systems fielen wiederkehrende Aufgaben an.

In regelmäßigen Abständen sollte zur Sicherheit ein Backup des gesamten Systems erfolgen. Das bei uns wöchentlich im Wechsel auf 2 Bändern durchgeführte Backup erfolgte automatisch, so daß nur das Wechseln der Bänder und deren Beschriftung und sichere Ablage anfiel. Diese regelmäßigen Aufgaben erforderten ca. 10 min pro Woche.

Der Vokabelmanager sollte mindestens einmal wöchentlich zur Abarbeitung der neuen Worte genutzt werden. Der Zeitrahmen wechselte in Abhängigkeit von der Anzahl der neuen Worte. Die Pflege des Lexikons erforderte ca. 15 min Arbeitszeit pro Woche.

Im dauerhaften Betrieb sollten alle 2- 3 Wochen die angefallenen Diktatzeiten kontrolliert und eventuell eine Adaptation durchgeführt werden. Der Zeitaufwand zur Bedienung des Systems betrug dafür ca. 30 min. Bei der akustischen Adaptation war der Spracherkennungsserver in Abhängigkeit von der angefallenen Diktatzeit zwischen 4 - 16 h pro Benutzer blockiert. Die Vokabelaktualisierung, die auf jedem Diktat-PC oder auf dem Spracherkennungsserver gestartet werden konnte, blockierte den jeweiligen Rechner pro Nutzer für ca. 4 h.

### 4.4 Entwicklung der Fehlerrate

#### 4.4.1 Fehlerrate nach dem Initialtraining

Die erste Adaptation wurde nach jeweils 9 Arbeitstagen durchgeführt, da sich innerhalb dieser Zeit genügend Sprachsignal angesammelt hatte. Daher wurde hier der Zeitraum dieser ersten 9 Tage betrachtet. Aufgrund der individuell unterschiedlichen Verteilung der befundeten Untersuchungen (Diagramm 1, Seite 28) wurden die jeweils am häufigsten befundenen Untersuchungen extra beurteilt. Das bedeutet, bei Arzt 1 und 3 wurden die Thoraxuntersuchungen, bei Arzt 2 die Neuro-CT extra beurteilt. Die durchschnittliche Fehlerrate nach dem Initialtraining und vor der ersten Adaptation betrug 8,4 - 13,3 % (Tabelle 4).

	Arzt 1		Arzt 2		Arzt 3	
	gesamt	Thorax	gesamt	CT-Neuro	gesamt	Thorax
Mittelwert (gesamt)	8,4 %	8,0 %	13,3 %	12,6 %	10,6 %	10,2 %
Standardabweichung	4,32	4,38	5,87	5,63	6,01	5,19

Tabelle 4 : Fehlerrate nach dem Initialtraining

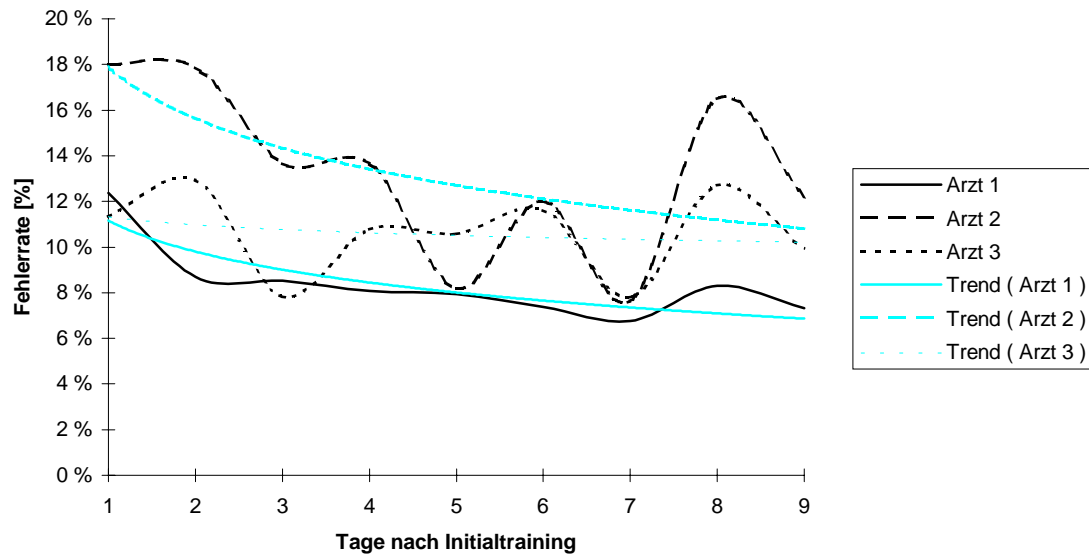


Diagramm 2 : Entwicklung der Fehlerrate nach dem Initialtraining

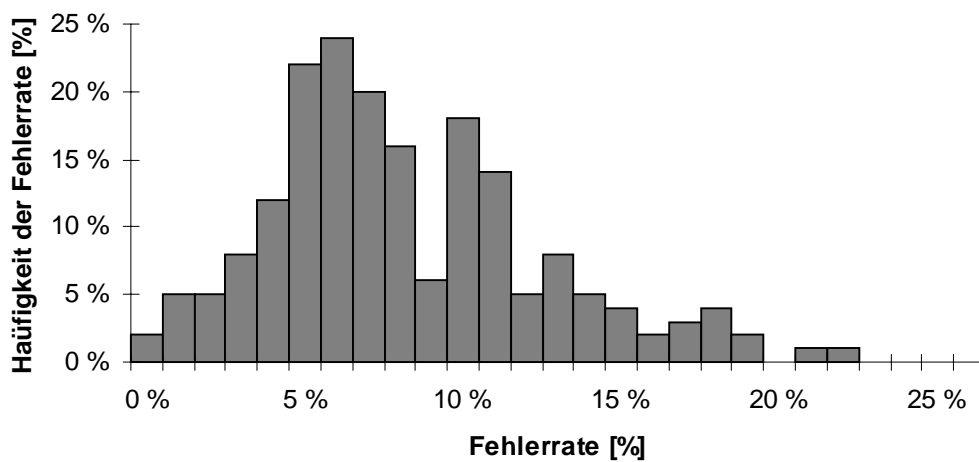
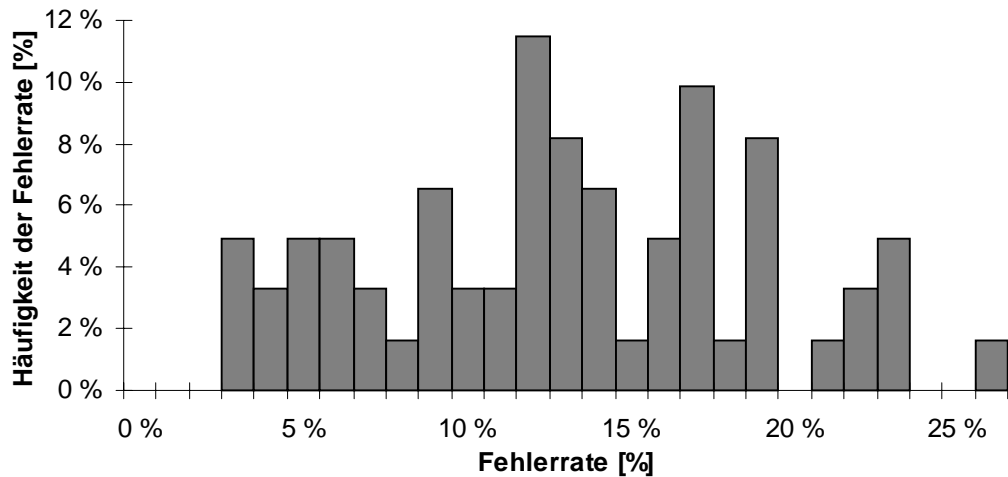
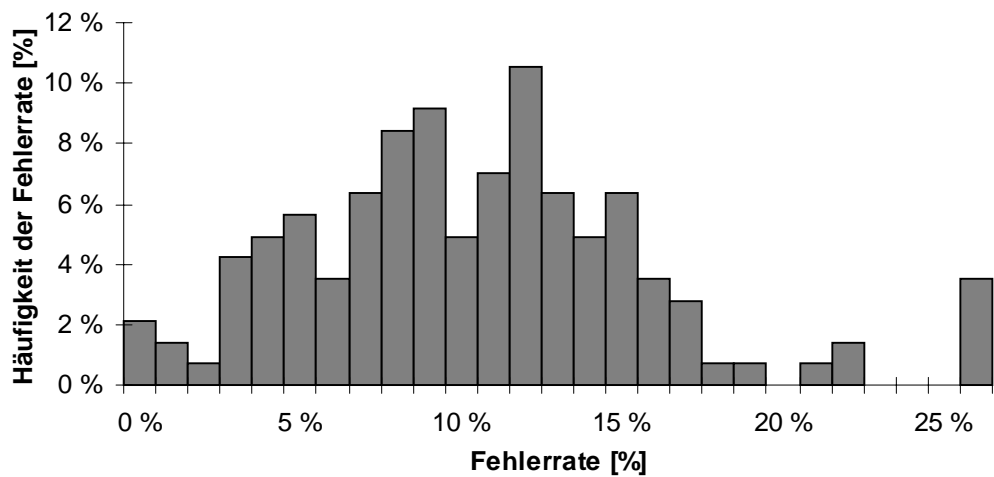


Diagramm 3 : Häufigkeitsverteilung der Fehlerrate nach dem Initialtraining (Arzt 1)



**Diagramm 4 : Häufigkeitsverteilung der Fehlerrate nach dem Initialtraining (Arzt 2)**



**Diagramm 5 : Häufigkeitsverteilung der Fehlerrate nach dem Initialtraining (Arzt 3)**

Im Verlauf (Diagramm 2, Seite 33) und der Häufigkeitsverteilung der Fehlerraten (Diagramm 3 - Diagramm 5) sowie an Hand der Standardabweichung ließen sich interindividuelle Unterschiede in der Schwankungsbreite feststellen.

Die im Diagramm 2 eingefügten Trendkurven veranschaulichen eine mehr oder weniger deutliche Verringerung der Fehlerrate über den Zeitraum zwischen Initialtraining und erster Adaptation, obwohl in diesem Zeitraum weder eine akustische Adaptation noch ein Vokaltraining durchgeführt wurde.

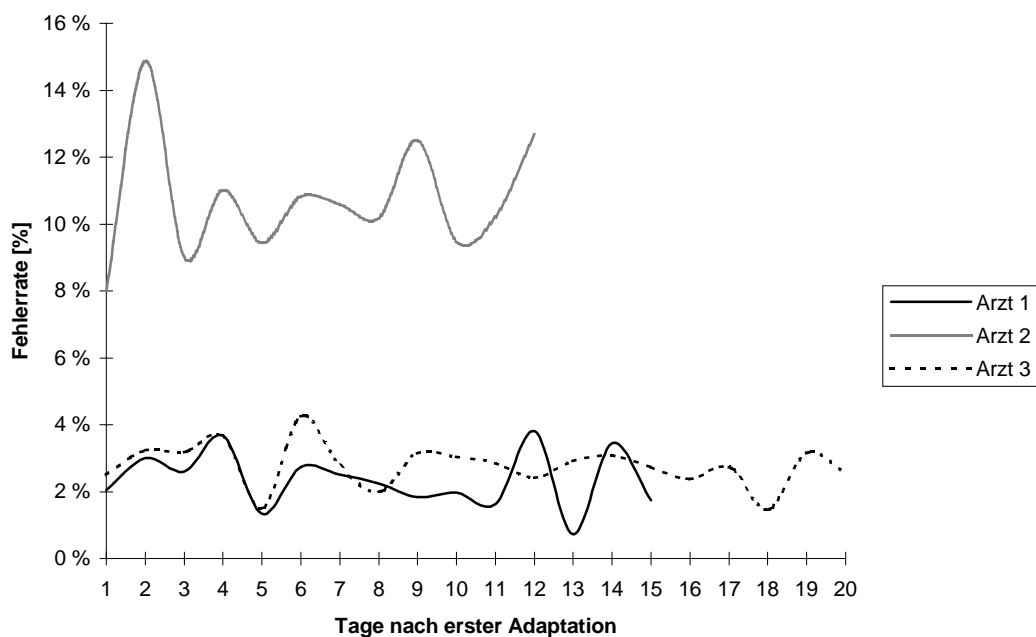
#### 4.4.2 Fehlerrate nach der ersten Adaptation

Der betrachtete Zeitraum liegt in der Zeit zwischen der ersten und der zweiten Adaptation. Da es durch den internen Arbeitsablauf und den Rotationsbetrieb zu unterschiedlichen Aufgabenverteilungen kam, konnte die zweite Adaptation nicht bei allen Testpersonen nach der gleichen Anzahl von Tagen erfolgen, so daß der hier betrachtete Abschnitt unterschiedlich lang ist, jedoch vergleichbar viel Sprachmaterial enthält (> 3 h Sprachsignal).

Nach der Durchführung der ersten Adaptation (akustisches Training und Vokabeltraining) kam es zu einer deutlichen Verbesserung der durchschnittlichen Erkennungsrate.

	Arzt 1		Arzt 2		Arzt 3	
	gesamt	Thorax	gesamt	CT-Neuro	gesamt	Thorax
Mittelwert	2,4 %	2,2 %	10,7 %	10,6 %	2,8 %	2,8 %
Standardabweichung	2,84	2,56	5,06	5,05	3,03	3,03

**Tabelle 5 : Fehlerrate nach erster Adaptation**

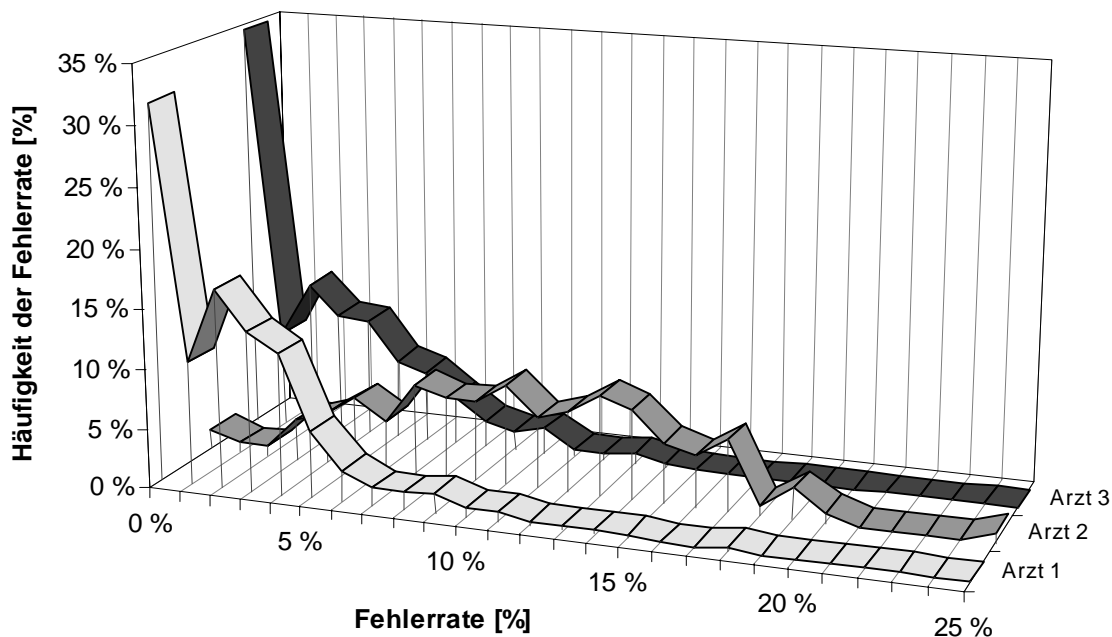


**Diagramm 6 : Entwicklung der Fehlerrate nach der ersten Adaptation**

Nach der ersten Adaptation lagen die durchschnittlichen Fehlerraten bei 2,4 - 10,7 % (Tabelle 5). Im Vergleich mit den durchschnittlichen Fehlerraten vor der ersten Adaptation (8,4 - 13,3 %; Tabelle 4, Seite 33) besteht eine deutliche Verringerung der Fehlerraten. Die Standardabweichung zeigte eine deutliche Regredienz. Die Signifikanzprüfung der Fehler-

raten vor und nach der 1. Adaptation (Thoraxdiagnostik Arzt 1:  $p < 0,01$ ; Arzt 3:  $p < 0,01$ ; CT-Neuro Arzt 2:  $p = 0,036$ ) bestätigte eine signifikante Verringerung der Fehlerraten bei Arzt 1, Arzt 2 und Arzt 3.

Die Häufigkeitsverteilung der Fehlerrate (Diagramm 7) zeigt, daß bei etwa 1/3 der erkannten Diktate bei Arzt 1 und Arzt 3 keine Fehler auftraten. Arzt 2 wies eine erhöhte Häufigkeit der Fehlerraten 7 - 17 % auf.



**Diagramm 7 : Häufigkeitsverteilung der Fehlerraten nach der ersten Adaptation**

Vergleicht man die Häufigkeitsverteilung der Fehlerraten nach der ersten Adaptation (Diagramm 7) mit der für den Zeitraum Initialtraining - 1. Adaptation (Diagramm 3 - Diagramm 5, Seite 33- 34), so ist vor allem bei Arzt 1 und Arzt 3, geringer auch bei Arzt 2 eine Verschiebung des Häufigkeitsmaximums in Richtung einer geringeren Fehlerrate zu erkennen.

### 4.4.3 Fehlerrate nach der zweiten Adaptation

Die Entwicklung der Fehlerrate nach der 2. Adaptation kann nur für Arzt 1 und Arzt 3 beurteilt werden. Arzt 2 nahm erst im späten Verlauf der Datenerfassung an der Testung des Spracherkennungssystems teil, so daß bis zur Beendigung der Datenerfassung mit ihm lediglich das Initialtraining und die 1. Adaptation durchgeführt werden konnten. Der betrachtete Zeitraum zwischen zweiter und dritter Adaptation bei den verbleibenden

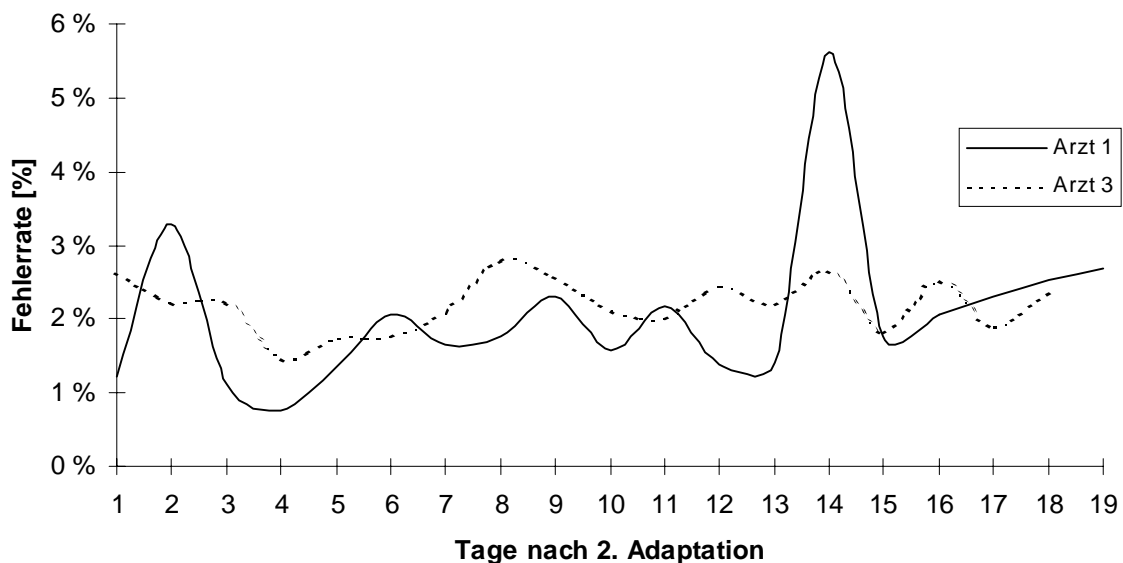
Testpersonen (Arzt1 und Arzt 3) war aus arbeitsorganisatorischen Gründen ebenfalls unterschiedlich groß.

	Arzt 1		Arzt 3	
	gesamt	Thorax	gesamt	Thorax
Mittelwert	2,1 %	1,7 %	2,2 %	2,2 %
Standardabweichung	2,34	1,91	2,41	2,39

**Tabelle 6 : Fehlerrate nach zweiter Adaptation**

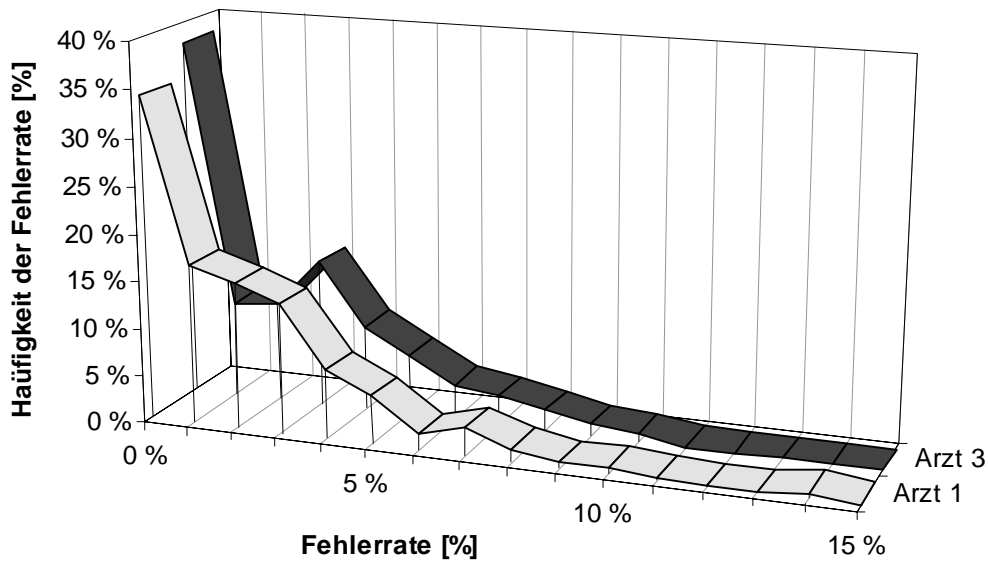
Im Vergleich zu den Fehlerraten nach der 1. Adaptation (2,4 % und 2,8 %; Tabelle 5, Seite 35) zeigte sich wiederum eine Verringerung der Fehlerrate auf 2,1 % und 2,2 % (Tabelle 6, Seite 37) sowie eine Verringerung der Standardabweichung bei beiden Nutzern. Diese Verbesserung fiel jedoch weitaus geringer aus als bei der 1. Adaptation.

Die Signifikanzprüfung der Fehlerraten vor und nach der 2. Adaptation (Thoraxdiagnostik Arzt 1:  $p < 0,01$ ; Arzt 3:  $p = 0,01$ ) beweist, daß es sich auch hier um eine statistisch signifikante, wenn auch geringe Verbesserung der Erkennungsrate handelte.



**Diagramm 8 : Entwicklung der Fehlerrate nach der zweiten Adaptation**

Der auffällige Anstieg der Fehlerrate bei Arzt 1 am Tag 14 nach der 2. Adaptation (Diagramm 8) fiel mit einem eintägigem Wechsel des Untersuchungsplatzes (Wechsel von der Thorax- zur Skelettbefundung ) zusammen.



**Diagramm 9 : Häufigkeitsverteilung der Fehlerraten nach der zweiten Adaptation**

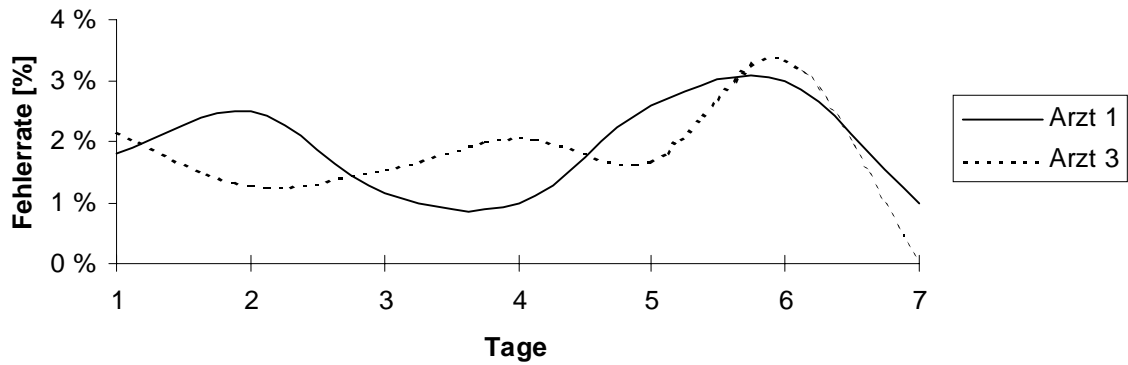
Der Vergleich der Häufigkeitsverteilungen der Fehlerraten vor und nach der 2. Adaptation (Diagramm 7; Seite 38 und Diagramm 9; Seite 41) zeigte eine geringe Zunahme der fehlerfrei bzw. fehlerarm erkannten Befunddiktate (Fehlerrate 0 - 3 %).

#### 4.4.4 Weitere Entwicklung der Fehlerrate (nach der dritten Adaptation)

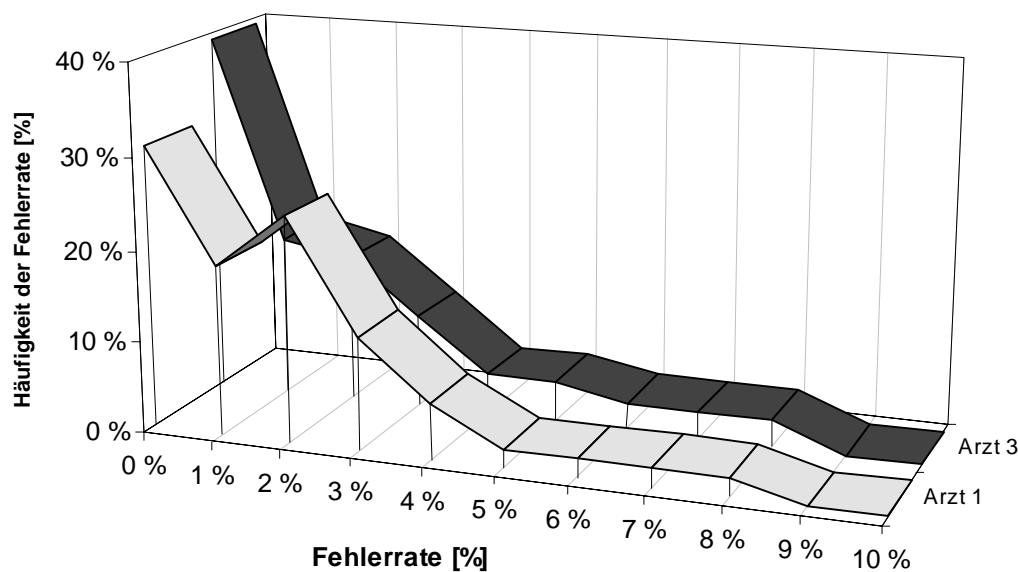
Sowohl bei Arzt 1 als auch bei Arzt 3 kam es im Vergleich zu dem Zustand vor der 3. Adaptation (Tabelle 6, Seite 37) zu einer Erhöhung der durchschnittlichen Fehlerrate (von 2,1 % auf 3,3 % bzw. von 2,2 % auf 3,1 %). Auf Grund der hausinternen Rotation wurden im Verlauf der Datenerfassung nach der dritten Adaptation verstärkt Befunde unterschiedlichster Untersuchungen erstellt.

	Arzt 1		Arzt 3	
	gesamt	Thorax	gesamt	Thorax
Mittelwert	3,3 %	1,9 %	3,1 %	1,8
Standardabweichung	2,67	2,32	2,19	2,16

**Tabelle 7 : Fehlerrate nach der dritten Adaptation**



**Diagramm 10 : Entwicklung der Fehlerrate nach der dritten Adaptation**



**Diagramm 11 : Häufigkeitsverteilung der Fehlerraten nach der dritten Adaptation**

Betrachtet man lediglich die Fehlerrate der Thoraxbefundungen (Tabelle 7, Seite 38) im Vergleich zu der Fehlerrate nach der 2. Adaptation (Tabelle 6, Seite 37), ergibt sich eine geringe Verbesserung der Erkennungsrate bei Arzt 3 (von 2,2 % auf 1,8 %) sowie ein geringer Anstieg der Fehlerrate von Arzt 1 (von 1,7 % auf 1,9 %) ohne eine wesentliche Veränderung des Verlaufs der Fehlerrate (Diagramm 10). Auch die Häufigkeitsverteilung der Fehlerraten (Diagramm 11) zeigte keine auffällige Veränderung im Vergleich zu den Daten nach der zweiten Adaptation (Diagramm 9, siehe Seite 38).

Die Signifikanzprüfung der Fehlerraten vor und nach der 3. Adaptation (Thoraxdiagnostik Arzt 1:  $p=0,49$  ; Arzt 3:  $p=0,26$ ) zeigte, daß die Veränderung der Fehlerrate der Thoraxbefundungen nach der 3. Adaptation bei Arzt 1 und Arzt 3 nicht signifikant war.

### 4.4.5 Fehlerrate in Abhängigkeit vom Individuum

Da von Arzt 1 und Arzt 3 eine genügend große Anzahl von Thoraxuntersuchungen befundet worden war, bot sich an, die individuellen Fehlerraten in Abhängigkeit vom Adaptationsstatus zu untersuchen. Um eventuell signifikante Unterschiede zu erkennen, wurde der ungepaarte Wilcoxon-Rangsummen-Test angewandt. Tabelle 8 zeigt die Entwicklung der Fehlerraten bei der Thoraxbefundung. Schon nach der ersten Adaptation traten interindividuell keine signifikanten Unterschiede in der Häufigkeit von falsch erkannten Worten auf ( $P=0,11$ ). Nach der 2. Adaptation stieg der Unterschied der individuellen Fehlerraten der beiden Ärzte wieder gering an ( $P=0,026$ ), nach der 3. Adaptation jedoch gab es erneut keinen signifikanten Unterschied ( $P=0,57$ ).

	Arzt 1 Fehlerrate [%]	Arzt 3 Fehlerrate [%]	Wahrscheinlichkeit p
Initialtraining	8,0	10,2	0,0
1. Adaptation	2,2	2,8	0,108
2. Adaptation	1,7	2,2	0,0263
3. Adaptation	1,9	1,8	0,5733

**Tabelle 8: Fehlerrate der Thoraxdiktate (Arzt 1 + Arzt 3)**

### 4.4.6 Fehlerrate in Abhängigkeit von der Untersuchungsart und -anzahl

Das Lexikon des SP 6000 enthielt am Anfang der Testphase vor allem Vokabeln aus der konventionellen Röntgendiagnostik, insbesondere der Thoraxdiagnostik. Um die Einsetzbarkeit an verschiedenen Arbeitsplätzen beurteilen zu können, war es wichtig, die Entwicklung der Fehlerraten unter dem Gesichtspunkt der verschiedenen Untersuchungstechniken zu betrachten. Untersuchungen, von denen nur Daten eines Diktates pro Zyklus vorlagen, wurden nicht berücksichtigt.

Nach der ersten Adaptation war bei allen Untersuchungsarten die Fehlerrate gefallen. Bei den Diktaten der Thoraxdiagnostik kam es jedoch zu einer stärkeren Verbesserung als bei allen anderen Untersuchungsarten.

	Rö Thorax		Rö Skelett		Rö Urologie		CT	
	Fehlerrate	Diktat-anzahl	Fehlerrate	Diktat-anzahl	Fehlerrate	Diktat-anzahl	Fehlerrate	Diktat-anzahl
nach Initialtraining	8,0 %	169	10,7 %	6	6,2 %	5	11,6 %	5
nach 1. Adaptation	2,4 %	319	5,5 %	6	3,5 %	6	4,7 %	4
nach 2. Adaptation	1,7 %	326	5,4 %	22	3,8 %	13	4,4 %	11
nach 3. Adaptation	1,9 %	49		0	4,2 %	75	3,0 %	8

**Tabelle 9 : Entwicklung der Fehlerrate in Abhängigkeit von der Untersuchungsart und -anzahl (Arzt 1)**

	CT Neuro		MRT Neuro	
	Fehlerrate	Diktat-anzahl	Fehlerrate	Diktat-anzahl
nach Initialtraining	12,6 %	56	12,5 %	2
nach 1. Adaptation	10,5 %	106	9,5 %	8

**Tabelle 10 : Entwicklung der Fehlerrate in Abhängigkeit von der Untersuchungsart und -anzahl (Arzt 2)**

	Rö Thorax		Rö Skelett	
	Fehlerrate	Diktat-anzahl	Fehlerrate	Diktat-anzahl
nach Initialtraining	10,0 %	140	32,0 %	3
nach 1. Adaptation	2,8 %	372		0
nach 2. Adaptation	2,2 %	478		0
nach 3. Adaptation	1,8 %	69	10,2 %	23

**Tabelle 11 : Entwicklung der Fehlerrate in Abhängigkeit von der Untersuchungsart und -anzahl (Arzt 3)**

Während es bei Arzt 3 zu einer weiteren, wenn auch geringen Verbesserung der Erkennungsrate der Thoraxdiagnostik nach der 2. und 3. Adaptation kam, war bei Arzt 1 nach der 3. Adaptation ein Verringerung der Erkennungsrate der Diktate der Thoraxdiagnostik und urologischen Diagnostik zu bemerken.

4.4.7 Fehlerrate in Abhängigkeit von der Diktatdauer

	Rö Thorax		Rö Skelett		Rö Urologie		CT	
	Fehlerrate	Diktatlänge (Mittelwert)	Fehlerrate	Diktatlänge (Mittelwert)	Fehlerrate	Diktatlänge (Mittelwert)	Fehlerrate	Diktatlänge (Mittelwert)
nach Initialtraining	8,0 %	74 s	10,7 %	67 s	6,2 %	92 s	11,6 %	218 s
nach 1. Adaptation	2,4 %	69 s	5,5 %	85 s	3,5 %	48 s	4,7 %	205 s
nach 2. Adaptation	1,7 %	74 s	5,4 %	79 s	3,8 %	71 s	4,4 %	172 s
nach 3. Adaptation	1,9 %	76 s			4,2 %	78 s	3,0 %	178 s

**Tabelle 12 : Fehlerrate in Abhängigkeit von der durchschnittlichen Diktatlänge (Arzt 1)**

	CT Neuro		MRT Neuro	
	Fehlerrate	Diktatlänge (Mittelwert)	Fehlerrate	Diktatlänge (Mittelwert)
nach Initialtraining	12,6 %	108 s	12,5 %	147 s
nach 1. Adaptation	10,5 %	103 s	9,5 %	146 s

**Tabelle 13 : Fehlerrate in Abhängigkeit von der durchschnittlichen Diktatlänge (Arzt2)**

	Rö Thorax		Rö Skelett	
	Fehlerrate	Diktatlänge (Mittelwert)	Fehlerrate	Diktatlänge (Mittelwert)
nach Initialtraining	10,0 %	64 s	32,0 %	82 s
nach 1. Adaptation	2,8 %	64 s		
nach 2. Adaptation	2,2 %	66 s		
nach 3. Adaptation	1,8 %	72 s	10,2 %	64 s

**Tabelle 14 : Fehlerrate in Abhängigkeit von der durchschnittlichen Diktatlänge (Arzt 3)**

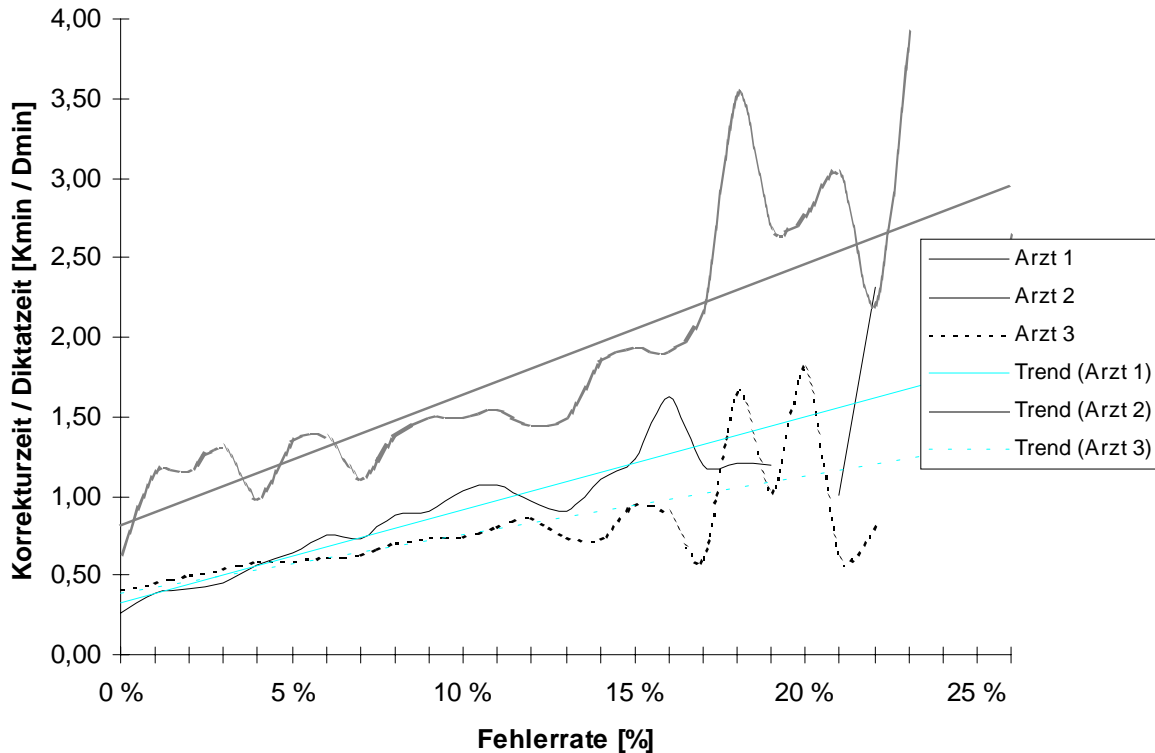
Es ließ sich keine Abhängigkeit der Fehlerrate von der durchschnittlichen Diktatlänge (Tabelle 12 - Tabelle 14) nachweisen.

	Arzt 1		Arzt 3	
	Fehlerrate (gesamt)	Diktatlänge (gesamt)	Fehlerrate (gesamt)	Diktatlänge (gesamt)
nach 1. Adaptation	2,4 %	246 min	2,8 %	151 min
nach 2. Adaptation	2,1 %	423 min	2,2 %	394 min
nach 3. Adaptation	3,3 %	436 min	3,1 %	474 min

**Tabelle 15 : Vergleich der Fehlerrate mit dem zugrundeliegenden Sprachsignal**

Bei dem Vergleich der Fehlerraten mit dem für die vorhergehende akustische Adaptation zur Verfügung stehenden Sprachsignal (Tabelle 15) war kein Zusammenhang erkennbar. Das bedeutet, daß eine Ansammlung von Sprachmaterial über die empfohlene Grenze von 2 h hinaus keine Verbesserung der Erkennungsleistung, aber eine Verlängerung der Dauer der akustischen Adaptation und damit der Blockierung des Spracherkennungsservers mit sich brachten.

#### 4.5 Kontrolle und Korrektur in Abhängigkeit von der Fehlerrate



**Diagramm 12 : Abhängigkeit des Korrekturaufwandes von der Fehlerrate**

Zur Beurteilung des Aufwandes für die Kontrolle und Korrektur des erkannten Textes wurde die Korrekturzeit erfaßt. Da der Zeitaufwand zum Lesen eines Textes in Abhängigkeit von der Textgröße (Diktatlänge) variierte, wurde zur Beurteilung des Korrekturaufwandes ein Quotient aus der Korrekturzeit und der Diktatlänge ( $Q(K,D)$ ) errechnet. Als Einheit wurde eine Korrekturminute pro Diktatminute gewählt ( $K_{\min} \cdot D_{\min}^{-1}$ ).

Wie durch die in das Diagramm 12 (Seite 42) eingefügten Trendkurven hervorgehoben wird, kam es erwartungsgemäß zu einer Verlängerung der Korrekturzeit bei verringerter Erkennungsrate (bzw. hoher Fehlerrate). Bei Fehlerraten von 17 % und mehr kam es zu einer Unregelmäßigkeit des Kurvenverlaufs.

Interessant ist auch der interindividuell unterschiedliche Kurvenverlauf. Hier wurde deutlich, daß die unterschiedliche Handhabung des Computers sowie die Schreibgeschwindigkeit einen nicht zu unterschätzenden Einfluß auf die notwendige Zeit zur Korrektur hatten.

### **4.6 Beurteilung der Befundungszeit**

Bei der Beurteilung der notwendigen Zeit zur Erstellung eines schriftlichen Befundes wurde nur die Arbeitszeit berücksichtigt, die sich bei den verschiedenen Befunderstellungsmethoden unterschied. Die Zeit zum Aufrufen der aktuellen Untersuchung des jeweiligen Patienten aus dem RIS, die Zeit der eigentlichen Befundung (Aufhängen der Bilder, Betrachtung und Beurteilung) und die Zeit zum Ausdruck des fertigen Befundes wurden nicht berücksichtigt.

Bei der Beurteilung der Befundungszeit wurden Untersuchungsarten, bei denen nur Daten zu einem Befunddiktat vorlag, außer acht gelassen.

#### 4.6.1 Zeitdauer zur Befunderstellung mittels SP 6000

Bei der Betrachtung der notwendigen Zeit zur Erstellung eines schriftlichen Befundes unter Nutzung des Spracherkennungssystems fiel als variable Größe die Diktatzeit sowie die notwendige Korrekturzeit ins Gewicht.

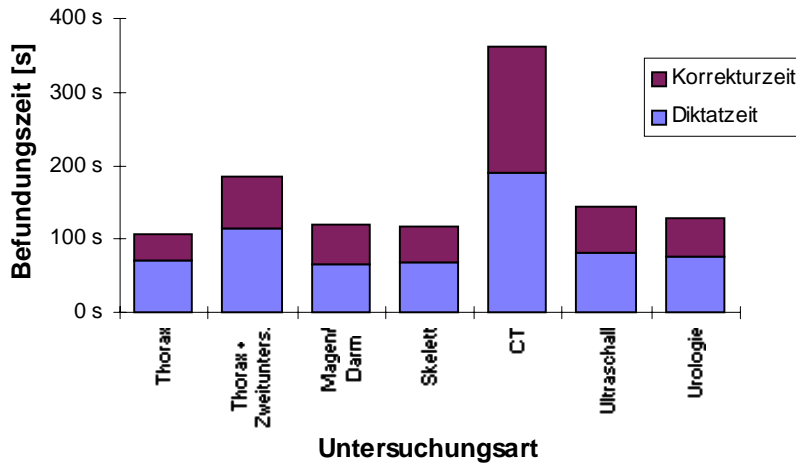


Diagramm 13 : durchschnittliche Befundungszeit (Arzt 1)

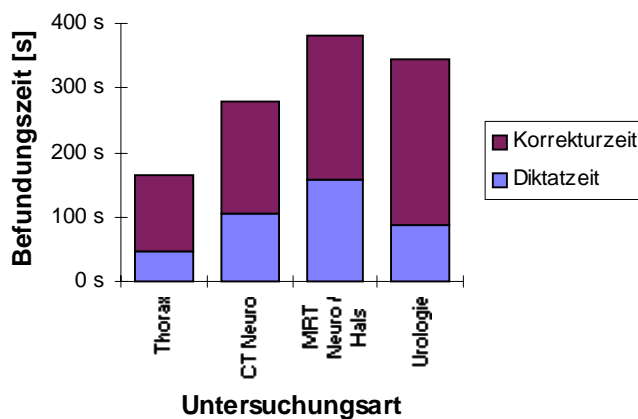


Diagramm 14 : durchschnittliche Befundungszeit (Arzt 2)

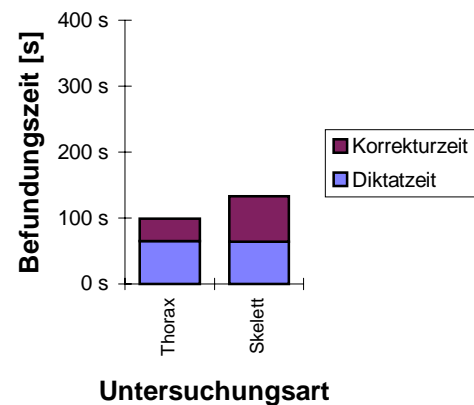


Diagramm 15 : durchschnittliche Befundungszeit (Arzt 3)

Die durchschnittliche Befundungszeit pro Diktat variierte zwischen 99 s und 380 s (Tabelle 16, Seite 45). Dabei spielten zum einen die Komplexität der Untersuchung (CT / MRT-Befundung komplexer als konventionelles Rö-Thorax), zum anderen die individuelle Fertigkeit bei der Bedienung des Korrektureditors eine große Rolle. Die relativ hohen Standardabweichungen (Tabelle 16) bei den verschiedenen Untersuchungsarten weisen auf die große Variabilität der Befundlänge innerhalb einer Untersuchungstechnik hin.

	Untersuchungsart	Diktatzeit	Korrekturzeit	Standardabweichung (Diktatzeit)	Standardabweichung (Korrekturzeit)
Arzt1	Thorax	71 s	35 s	19,27	24,99
	Thorax + Zweitunters.	115 s	71 s	20,02	26,57
	Magen/ Darm	66 s	54 s	7,00	16,77
	Skelett	67 s	50 s	35,10	43,75
	CT	190 s	171 s	64,23	177,83
	Ultraschall	83 s	62 s	24,47	26,08
	Urologie	76 s	53 s	30,19	31,37
Arzt 2	Thorax	45 s	120 s	18,38	84,85
	CT Neuro	105 s	176 s	34,02	110,07
	MRT Neuro / Hals	159 s	221 s	45,26	97,58
	Urologie	86 s	260 s	13,32	69,28
Arzt 3	Thorax	65 s	34 s	20,00	14,13
	Skelett	64 s	69 s	31,03	58,36

**Tabelle 16 : mittlere Diktatzeit, mittlere Korrekturzeit und Standardabweichung für beide Zeiten**

#### 4.6.2 Zeitdauer zur Befunderstellung mittels Schreiben

Tabelle 17 (Seite 47) zeigt die durchschnittliche Schreibdauer für Befunde der unterschiedlichen Untersuchungstechniken. Unterschieden wurde zwischen Schreiben mit und ohne Textbaustein.

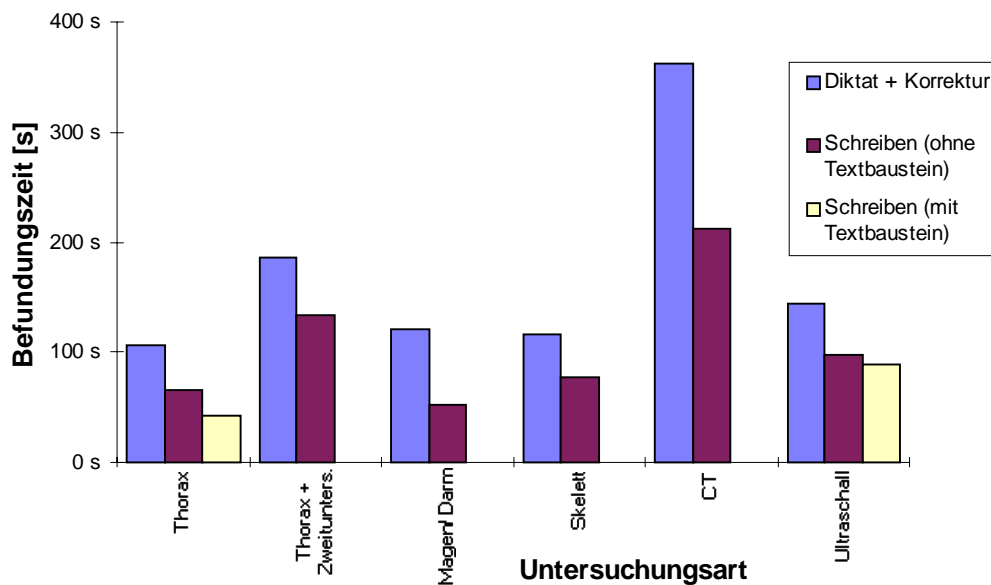
Erwartungsgemäß brachte die Nutzung von Textbausteinen eine erhebliche Zeiteinsparung. So konnte zum Beispiel in der Thoraxdiagnostik die Schreibdauer von Arzt 1 von 119 s auf 42 s unter Nutzung von Textbausteinen gesenkt werden. Ähnlich der Befunderstellung über das Diktieren (Tabelle 16, Seite 46) zeigte sich auch bei der schriftlichen Befundung eine hohe Standardabweichung der Befunderstellungsdauer.

		Befundart	Anzahl der Befunde	Schreibdauer	Standardabweichung
Arzt 1	ohne Textbaustein	Rö Thorax	119	66 s	29,32
		Thorax + Zweitunters.	6	133 s	10,41
		Rö Skelett	9	78 s	53,11
		Rö Trauma	181	31 s	28,50
		Rö Abdom	3	16 s	5,13
		Rö Magen/Darm	2	53 s	31,82
		US Abdom	106	105 s	69,17
		US Abdom + Zweitunters.	23	163 s	127,90
		US Niere	5	87 s	22,93
		Rö HNO	5	40 s	21,87
		US FKDS Hals/Bein	8	84 s	33,48
		US Thorax	5	31 s	9,74
		US Weichteile / Mamma	3	90 s	39,31
		US Hals	5	85,8	23,61
		CT	2	212,5	123,74
	mit Textbaustein	Rö Thorax	22	42 s	21,89
		Rö Trauma	82	8 s	10,66
		US Abdom	81	82 s	30,99
		US Abdom + Zweitunters.	20	115 s	111,51
Arzt 3	ohne Textbaustein	Rö Thorax	115	156 s	28,28
	mit Textbaustein	Rö Thorax	27	18 s	4,94

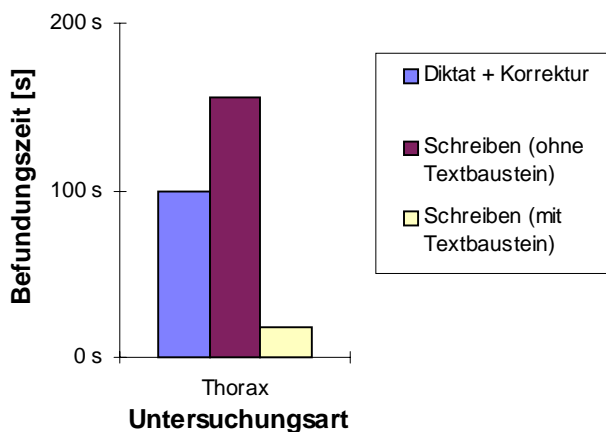
Tabelle 17 : Schreibdauer ( Arzt 1 + Arzt 3)

### 4.6.3 Vergleich der Befundungszeit des Schreibens mit der des Diktierens

Zum Vergleich der Befundungszeit wurden nur die Untersuchungsarten berücksichtigt, für die bei beiden Befundungstechniken Daten erfasst worden waren.



**Diagramm 16 : Befundungszeiten Diktat / Schreiben (Arzt 1)**



**Diagramm 17 : Befundungszeiten Diktat / Schreiben (Arzt 3)**

Da bei den Daten der Diktate die Ultraschalluntersuchungen nicht einzeln klassifiziert wurden, wurde als Vergleichswert des Schreibens der Mittelwert aller US-Befundungen gebildet.

Diagramm 16 (Seite 48) zeigt bei Arzt 1 deutlich den Geschwindigkeitsvorteil des manuellen Schreibens gegenüber der Befundungszeit beim Diktieren. Bei einem Zeitvergleich (Tabelle 16, Seite 46, Tabelle 17, Seite 47) am Beispiel der Thoraxdiagnostik (Diktat/Korrektur:

106 s; Schreiben: 66 s; Schreiben mit Textbaustein: 42 s) konnte Arzt 1 eine Zeitersparnis von ca. 40 % durch das Schreiben erreichen.

Diagramm 17 (Seite 48) zeigt hingegen einen deutlichen Zeitvorteil des Diktierens gegenüber dem Schreiben bei Arzt 3. Hier erbrachte der Vergleich der Befundungszeiten (Tabelle 16, Seite 46, Tabelle 17, Seite 47) am Beispiel der Thoraxdiagnostik (Diktat/Korrektur: 99 s; Schreiben: 156 s; Schreiben mit Textbaustein: 18 s) einen Zeitvorteil des Diktierens von ca. 40 % gegenüber der herkömmlichen Schreibweise.

In jedem Fall brachte die Nutzung von Textbausteinen einen nicht unerheblichen Zeitgewinn bei der schriftlichen Befunderstellung.

## 5 Diskussion

### 5.1 Methodik

#### 5.1.1 Testpersonen

Die Testpersonen wurden willkürlich gewählt. Da das SP 6000 zu Beginn unserer Testung noch kein kommerzielles Produkt darstellte und wir mit einer Testversion arbeiteten, wurden die Ärzte 1 und 3 wegen ihrer Computerkenntnisse ausgewählt, um bei eventuellen Schwierigkeiten (Programmabstürzen, Fehlermeldungen,...) reagieren und trotzdem im Routinebetrieb mitarbeiten zu können. Arzt 2 wurde aus Eigeninteresse später einbezogen. Bei der Bewertung der individuell unterschiedlichen Datenmengen und Ergebnisse muß der jeweilige Ausbildungsstand, die Diktiererfahrung und die Tastaturfertigkeit des Einzelnen berücksichtigt werden.

Die Testpersonen ergeben somit keinen repräsentativen Querschnitt aller Ärzte aus dem Bereich der radiologischen Diagnostik.

#### 5.1.2 Meßwerte

Bei der Interpretation der Ergebnisse sind folgende Randbedingungen zu berücksichtigen:

- Es wurden zu den einzelnen Untersuchungsarten nicht in jedem Fall genügend Daten erfaßt ( z.B. Ultraschallbefundung nur 8 Diktate, Tabelle 1; Seite 27).
- Es liegen keine Informationen zu der Komplexität der einzelnen Untersuchungen vor (pathologisch, Normalbefund etc.), um die große Schwankungsbreite der Diktatlängen sowie extreme Fehlerraten erklären zu können
- Bei der Errechnung der Fehlerrate berücksichtigte der Korrektureditor auch versehentlich geänderte Worte oder Zahlen (bei denen die versehentliche Änderung zurückgenommen wurde) als falsch erkannte Worte.
- Es wurden keine Aufzeichnungen über eventuelle Änderungen der Phonetik (Schnupfen, Heiserkeit, Zahnbehandlungen,...) der Testpersonen, die eine kurzfristige Änderung der Erkennungsrate zur Folge gehabt haben könnten, vorgenommen.
- Bei der Zeitmessung der schriftlichen Befundung mußte der gesamte Befund en bloc geschrieben werden. Als Alternative wurde bei Denkpausen die Stopuhr angehalten und erst beim Weiterschreiben wieder gestartet. Nur so konnte gewährleistet sein, auch die tatsächliche Schreibzeit zu erfassen.

- Die Daten der schriftlichen Befundung konnten nur von 2 Personen erfaßt werden. Deren Schreibgeschwindigkeit läßt sich jedoch nicht ohne weiteres verallgemeinern (Arzt 1 sehr schnell, Arzt 2 langsamer).
- Die schriftliche Befundung ist in der Regel in der Befundbeschreibung und den klinischen Angaben kürzer gefaßt als der diktierte Befund und konzentriert sich direkt auf das Untersuchungsergebnis. Das bedeutet, daß die absolute Textinformation eines geschriebenen Diktates oft weniger beinhaltet als ein diktiert Befund.

## **5.2 Entwicklung der Fehlerrate**

Die durchschnittliche Fehlerrate zeigte schon nach dem Initialtraining Werte (Tabelle 4, Seite 33) zwischen 8,4 - 13,3 % (synonym Erkennungsrate 86,7 - 91,6 %), das bedeutet ein falsch erkanntes Wort auf etwa 10 Worte bzw. einen Satz. Schon innerhalb der ersten Tage nach dem Initialtraining verbesserte sich die Erkennungsrate gering (Diagramm 2, Seite 33). Da in diesem Zeitraum keine Veränderungen am Spracherkennungssystem vorgenommen wurden, kann es sich in diesem Fall nur um eine „Adaptation“ der Benutzer an das Spracherkennungssystem handeln. Diese „Adaptation“ bedeutet einen sichereren Umgang mit dem System (Mikrofon, Korrekturator), ein Erlernen des Diktierens mit Interpunktion, Formatierung und Steuerworten, Vermeidung des „Verschluckens“ von Endungen, etc.

Das System zwang den einzelnen Nutzer zu einer zum Teil veränderten Sprechweise. Dabei mußte keine Rücksicht auf den Wortfluß an sich genommen werden, da das SP 6000 als kontinuierliche Spracherkennung keine Pausen zwischen den Worten benötigt, eher dadurch gestört wird. Die veränderte Sprechweise bezieht sich vielmehr auf die korrekte Aussprache. Das Lexikon enthält außer den Worten an sich auch die entsprechenden Phoneme, aus denen das Wort aufgebaut ist. Bei der Texterkennung werden die Phoneme des Diktats mit den im Lexikon vorhandenen Phonemen verglichen. Bei einer unsauberen Aussprache erhält das System keine korrekten Phoneme zur Beurteilung und kann folglich nicht das korrekte Wort erkennen. Die Folge ist ein ständiges Korrigieren dieses einen Wortes. Die einzige Möglichkeit sich diese lästige Tätigkeit zu vereinfachen, ist die korrekte Aussprache.

Die erhebliche Verbesserung der durchschnittlichen Erkennungsrate (Tabelle 5, Seite 35) nach der ersten Adaptation auf Werte von 2,4 %, 10,7 % und 2,8 % (bei Arzt 1, Arzt 2 und Arzt 3) beweist, daß die Phonetik des freien Diktierens sich erheblich von der des Vorlesens unterscheidet. Die geringe Veränderung der durchschnittlichen Fehlerrate nach der 2. und 3. Adaptation (Tabelle 6, Seite 37 und Tabelle 7, Seite 38) auf Werte zwischen 2,1 % - 3,3 % weist auf die geringe Variabilität der Phonetik des Diktierens hin. Diese Verände-

rungen sind eher auf die Erweiterung des Lexikons und auf Anpassungen im persönlichen Sprachmodell zurückzuführen.

Der geringe Anstieg der Fehlerrate nach der 3. Adaptation hängt in hohem Maße mit dem Wechsel des Befundarbeitsplatzes zusammen. Durch das Auftreten neuer, dem Lexikon unbekannter Vokabeln vermehrte sich das Auftreten falsch erkannter Worte, die erst nach der nächsten Vokabelaktualisierung erkannt werden können.

Betrachtet man isoliert die Entwicklung der Fehlerrate bei Thoraxbefunden, so bestätigt das Fehlen signifikanter Unterschiede (siehe 4.4.4, Seite 38) nach der dritten Adaptation diese Vermutung. Da sich das Vokabular und die Sprechweise in diesem Arbeitsbereich nicht änderten, kam es auch zu keinem signifikanten Unterschied der Erkennungsrate gegenüber der Erkennungsrate vor der 3. Adaptation.

Bei der Betrachtung der Fehlerrate unter dem Gesichtspunkt des einzelnen Befunders kommen wir zu dem Ergebnis, daß schon nach der ersten Adaptation eine interindividuelle Gleichverteilung der Fehlerrate zustande kommt (Tabelle 8, Seite 40). Nach der 2. Adaptation traten wieder signifikante Unterschiede der individuellen Fehlerraten auf. Diese Unterschiede zeigten jedoch in der praktischen Befundungstätigkeit keine Relevanz. Geht man von einer durchschnittlichen Satzlänge von 10 Worten aus, so mußte bei einer Fehlerrate (bei der Thoraxbefundung) von durchschnittlich 1,7 - 2,2 % in jedem 5. bis 6. Satz jeweils ein Wort korrigiert werden. Die Wahrscheinlichkeit von  $p=0,57$  nach der 3. Adaptation beweist wieder die Gleichverteilung der interindividuellen Fehlerrate. Die Verlaufsbeobachtung der interindividuellen Gleichverteilungüberprüfung läßt die Vermutung zu, daß sich bei Beachtung der Grundprinzipien der Nutzung des Spracherkennungssystems SP 6000 (siehe 3.1.1 - 3.1.6, Seite 21 - 24) schon nach wenigen Trainings- bzw. Adaptationsdurchläufen eine vom individuellen Befunder relativ unabhängige Erkennungsrate einstellt.

Vergleicht man die Entwicklung der Fehlerraten in Abhängigkeit von den zugrunde liegenden Untersuchungen (Tabelle 9 - Tabelle 11, Seite 41), so wird deutlich, daß sich die Erkennungsrate der Thoraxbefundungen höhergradig verbesserte als die der Untersuchungen mit geringerer Diktanzahl. Die Erklärung für die bessere Erkennungsrate ist zum einen in der großen Anzahl der Diktate zu suchen, da hier in kurzer Zeit das individuell genutzte Spektrum der verwendeten Worte durch die Adaptationen (mit Vokabelaktualisierung) fast komplett im Lexikon vorlag. Einen weiteren Anhaltspunkt bietet die konventionelle Thoraxbefundung an sich, da in den meisten Fällen ein begrenztes Sprachspektrum zur Befundung ausreicht. Ein weiteres Kriterium der Befundung in der Thoraxdiagnostik ist der relativ hohe Prozentsatz an unauffälligen Befunden, da viele Thorax-

aufnahmen z.B. zur Verlaufskontrolle, OP-Vorbereitung oder Vorsorge bei TBC-Kontakt gemacht werden, so daß relativ häufig ein bis auf die anamnestischen Angaben nahezu wortgleicher Befund erstellt wird.

Vergleicht man die hochgradige Verbesserung der Thoraxerkennungsrate nach der 1. Adaptation mit der nur geringen Verbesserung bei der Befundung von neurologischen CT-Untersuchungen (Tabelle 4, Seite 33), so bestätigt die geringe Verbesserung obige Ausführungen. Schon durch ausführlichere klinische Angaben zu jeder CT-Untersuchung wird das Spracherkennungssystem ständig mit Vokabeln aus dem nichtradiologischen Kontext konfrontiert, die bis zur nächsten Vokabelaktualisierung ständig korrigiert werden müssen. Durch die umfassende Befundung der komplexen Untersuchung wird ein weitaus größeres Spektrum an individuell unterschiedlichen Vokabeln genutzt, so daß das System relativ viele neue Vokabeln integrieren, gleichzeitig aber auch ein individuell komplexeres Sprachmodell erzeugt werden muß. Die Erzeugung eines komplexeren Sprachmodells setzt jedoch eine ausreichend große Datenmenge voraus. Leider konnte die (aufwendige) Datenerfassung bei der spracherkennungsgestützten Befundung von neurologischen CT-Untersuchungen nicht fortgesetzt werden, so daß hier der weitere Verlauf der Erkennungsrate nicht beurteilt werden kann.

Die akustische Adaptation sollte laut Herstellerangaben mit mindestens 2 h Sprachsignal erfolgen. Bei dem Versuch, die Erkennungsrate mit unterschiedlicher Diktatmenge für die akustische Adaptation (Tabelle 15, Seite 43) zu beeinflussen, konnte keine signifikante Verbesserung beobachtet werden.

### **5.3 Kontrolle und Korrektur**

Ähnlich dem konventionellen Diktiervorgang, bei dem der Arzt den geschriebenen Befund auf der Sekretärin unbekannte Worte „und Schreibfehler“ kontrolliert, ist auch bei der Spracherkennung die Kontrolle des Inhalts und der unbekanntem Worte des erkannten Textes unbedingt notwendig. Ein Vorteil der digitalen Spracherkennung ist, daß Rechtschreib- (oder Tipp-) fehler nicht vorkommen. Die erkannten Worte werden so geschrieben, wie sie in dem Lexikon eingetragen wurden. Es könnten nur fehlerhaft eingetragene Termini einen Rechtschreibfehler verursachen. Während unseres Testes trat dieser Fehler nicht auf. Das Augenmerk bei der Kontrolle ist daher auf falsch erkannte Worte zu legen. Es gibt verschiedene Kriterien, durch die ein Wort falsch erkannt wird:

- Fehlerhafte Aussprache, das heißt das Wort wird phonetisch nicht korrekt ausgesprochen. Wenn jedoch der Nutzer einen Sprachfehler (z.B. Lispeln) hat, so wird dieser

Fehler bei dem Initialtraining bzw. der akustischen Adaptation dem entsprechenden Phonem zugeordnet und entsprechend erkannt.

- Bei unsauberer Sprechweise, wie z.B. das „Verschlucken“ von Endungen bei Worten wie „einen“ und „einem“ ist das System überfordert und es treten Fehler in der Erkennung auf.
- Es gibt ein phonetisch sehr ähnliches Wort, das im Sprachmodell eine höhere Verwendungswahrscheinlichkeit hat.
- Das diktierte Wort ist noch nicht im Lexikon vorhanden. Es wird ein phonetisch ähnliches Wort bzw. eine Wortgruppe angeboten.

Die ersten beiden Fehlerursachen sind durch konsequente Aussprachedisziplin korrigierbar. Die letzten beiden Fehlerquellen können erst durch die nächste Vokabelaktualisierung beseitigt werden. Durch jede Korrektur erhält das verwendete bekannte Wort eine Erhöhung seiner Nutzungswahrscheinlichkeit. Jedes neue Wort wird mittels des Vokabelmanagers kontrolliert und eventuell als neuer Lexikoneintrag definiert. Da das SP 6000 kein kontinuierlich lernendes System ist, wird es passieren, daß in der Anfangsphase oder bei der Arbeit an einem neuen Befundarbeitsplatz bestimmte Fehler immer wieder auftauchen und korrigiert werden müssen. Das ständige Korrigieren hat aber zur Folge, daß die Wahrscheinlichkeit der Nutzung der korrigierten Worte steigt. Die veränderte Wahrscheinlichkeit wird bei der Vokabelaktualisierung im Sprachmodell des Nutzers eingetragen, so daß nach der Adaptation dieser Fehler nicht mehr auftauchen sollte.

Die dritte Fehlerursache stellt bei ungenauer Kontrolle eine große Gefahr für die korrekte Befundung dar. Es gibt verschiedene Worte, die sich phonetisch sehr ähnlich sind, aber eine unterschiedliche Aussage treffen. So klingen Worte wie „ein“ und „kein“ sehr ähnlich, ergeben aber eine gegensätzliche Aussage („eine“ oder „keine Metastase“). Außerdem gibt es in der Medizin viele Abkürzungen, die als Wort gesprochen werden. So fielen uns in der Skelettdiagnostik die „PIP-“ und „DIP-“Gelenke (proximales bzw. distales Interphalangealgelenk) auf. Diese (Kunst-) Worte werden in etwa gleich häufig benutzt und klingen sehr ähnlich. Auch bei korrekter Aussprache ist es für das Spracherkennungssystem sehr schwierig den jeweils richtigen Terminus zu erkennen.

Der Vergleich der notwendigen Korrekturzeit mit der dazugehörigen Fehlerrate (Diagramm 12, Seite 43) ergab erwartungsgemäß eine direkte Abhängigkeit. Interessanterweise stellte sich ein relativ großer Unterschied der Korrekturzeiten bei gleichen Fehlerraten der Testpersonen heraus. Dabei korreliert die notwendige Korrekturzeit in etwa mit der Einschätzung der Computer- und Tastaturfertigkeiten. Die Unregelmäßigkeit des Kurvenverlaufs bei

Fehlerraten von 17 % und mehr ist durch die wenigen schlecht erkannten Diktate zu erklären, da hier der mittlere  $Q(K/D)$  aus ein bis 2 Diktaten ermittelt wurde, und somit keine repräsentativen Aussagen erhalten werden konnten.

Bei der Netzwerkvariante des SP 6000 besteht die Möglichkeit, die erkannten Diktate an jedem Diktat-PC zu korrigieren. Der Befunder hat auch die Option, seine erkannten Diktate an einen bestimmten Nutzer zur Korrektur freizugeben. So könnte eine zentrale Korrektur-(Schreib-)kraft die Korrektur der erkannten Texte vornehmen.

Nach unseren Erfahrungen mit dem Spracherkennungssystem SP 6000 ist es jedoch anzuraten, daß wenigstens in der Anfangszeit jeder Nutzer die Korrekturen selber vornimmt, damit er die Möglichkeit hat, die eigene Sprache (Aussprache, Interpunktion, Steuerwortnutzung) zu kontrollieren und zu verbessern.

#### **5.4 Befundungszeit**

Der hier betrachtete Zeitraum schließt nur die unmittelbare Erstellung des schriftlichen Befundes ein. Die eigentliche Beurteilung der Untersuchung, deren Dauer in erster Linie vom Ausbildungsstand und der Erfahrung des Befunders abhängt, wurde dabei nicht berücksichtigt, wobei der Ausbildungsstand und die Erfahrung sich natürlich auch auf die Erstellung des Befundes auswirken. Nach dem Erkennen eines Befundes muß dieser noch formuliert werden. Auch dieses Formulieren muß erlernt werden und kann eine verlängernde Wirkung auf die schriftliche Befundung haben. Die von uns untersuchte Befundungsdauer schließt bei der digitalen Spracherkennung Diktat- und Korrekturzeit ein (Kapitel 4.6.1, Seite 44). Nicht in die Datenerfassung eingeflossen ist die Zeitdauer der eigentlichen Spracherkennung. Das SP 6000 benötigte zur Umwandlung des Sprachsignals in geschriebenen Text je nach Adaptationsgrad des Befunders ca. 1- 2 Minuten pro Diktatminute. Um nach Fertigstellung eines Diktates nicht auf die Beendigung der Spracherkennung dieses Diktates warten zu müssen, diktierten wir in der Regel 3- 5 Untersuchungen, um dann mit der Korrektur zu beginnen. Während des zweiten Diktates konnte die Erkennung des vorhergehenden im Hintergrund ablaufen. Wenn mit der Korrektur des ersten Diktates begonnen wurde, konnte im Hintergrund die Erkennung der letzten Diktate fertiggestellt werden. Auf diese Art und Weise konnten Leerlaufzeiten vermieden werden und wurden daher auch nicht erfaßt.

Außer der oben erwähnten interindividuellen Unterschiede der Korrekturzeit fällt die untersuchungsartbezogene unterschiedlich lange Befundungszeit (Diagramm 13, Seite 45 ) auf. Der relativ hohe Werte der Standardabweichung, der bei allen Untersuchungen zu finden

ist (Tabelle 16, Seite 46), beweist eine große Variabilität der Befundzeit innerhalb der einzelnen Untersuchungstechniken. Die Variabilität der notwendigen Befundungszeit im Rahmen einer Untersuchungsart kommt durch die große Bandbreite der Untersuchungsergebnisse zwischen normal bis hoch pathologisch und auch durch unterschiedlich komplexe Teiluntersuchungen zustande (z.B. bei der konventionellen Skelettdiagnostik: Verlaufsbeurteilung einer Endoprothese - Untersuchung beider Hände und Füße in der Rheumadiagnostik).

Der auffällige Zeitunterschied zwischen den einzelnen Untersuchungsarten kommt am deutlichsten bei dem Vergleich der Schnittbildverfahren mit der konventionellen Röntgen-diagnostik zum Ausdruck. Durch die Notwendigkeit der genauen Erfassung der Untersuchungsdurchführung, der detaillierten Befundbeschreibung und der letztendlichen Beurteilung ergeben sich komplexere Befunde der Schnittbildverfahren im Vergleich zur konventionellen Thoraxdiagnostik.

Um einen Vergleich der digitalen Spracherkennung mit der in unserem Institut in großem Umfang genutzten Art und Weise der Befunderstellung durchführen zu können, wurden Daten vom Schreiben der Befunde erhoben. Auch hier stellten sich große Unterschiede der Befundungszeit zwischen den einzelnen Untersuchungen heraus (Tabelle 17, Seite 47). Die ebenfalls beim Schreiben ermittelten hohen Standardabweichungen der Befundungszeit belegen eine große Bandbreite der Befundlängen innerhalb einer Untersuchungstechnik. Erwartungsgemäß kommt es bei dem Vergleich der individuell notwendigen Zeit bei der schriftlichen Befundung zu erheblichen Unterschieden.

Die Möglichkeit des durch Textbausteine unterstützten Schreibens wird auch in der Befundung eingesetzt. Die Daten der Befunde, die auf diese Art und Weise erstellt wurden, wurden gesondert gekennzeichnet. Im Vergleich mit dem Schreiben des kompletten Befundes kommt es zu einer zum Teil erheblichen Verkürzung der Schreibdauer. Die Verkürzung der Schreibdauer hängt von der individuellen Nutzung der Textbausteine ab. Werden nur komplette Befunde (z.B. ein unauffälliger Thorax) als Textbaustein genutzt, ergeben sich im Einzelfall sehr kurze Schreibzeiten. Wendet man jedoch Textbausteine für Worte und Wortgruppen an, verkürzt sich die Schreibzeit im Einzelfall weniger, dafür können diese Textbausteine aber auch bei Nichtstandard-Befunden, und damit häufiger eingesetzt werden.

In der uns vorliegenden Testversion des SP 6000 war die Nutzung von Textbausteinen nicht gegeben. In der Endversion soll eine Möglichkeit zum Einsatz von Textbausteinen implementiert werden. Allerdings ist hier die Editiermöglichkeit des Textes eines Textbausteines nicht während des Diktates gegeben. Erst nach Übergabe des korrigierten Textes

an das Befundungssystem können eventuelle Anpassungen des Textbausteins an den jeweiligen Befund vorgenommen werden.

Im Vergleich der Befundungszeiten Diktieren/Schreiben (Diagramm 16 und Diagramm 17, Seite 48) fällt bei Arzt 1 ein Zeitvorteil des Schreibens auf. Bei Arzt 3 jedoch tritt ein deutlicher Zeitvorteil des Diktierens und Korrigierens gegenüber der normalen schriftlichen Befundung auf. Bei der Beurteilung eines Zeitvorteils in dieser oder jener Richtung müssen in jedem Fall die persönlichen Fähigkeiten einbezogen werden. So schreibt Arzt 1 mit einer sehr hohen Schreibgeschwindigkeit, während Arzt 3 mit einer sehr guten, aber nicht professionellen Tastaturfertigkeit arbeitet. Bei ungeübten Tastaturnutzern dürfte der Zeitvorteil eindeutig auf Seiten der Spracherkennung liegen.

Bei der Interpretation der Ergebnisse des Vergleichs zwischen geschriebenen und diktierten Befunden muß auch berücksichtigt werden, daß sich nach unseren Erfahrungen die diktierten von den geschriebenen Befunden auch in ihrem Umfang unterscheiden. Die diktierten Befundungen enthalten z. B. die auf der Untersuchungsanforderung mitgeteilten klinischen und anamnestischen Daten. Geschriebene Befunde enthalten oft wenig oder keine zusätzlichen Informationen und sind auch in der Befundbeschreibung oft knapp gehalten. Sicher würde sich die Schreibzeit für einen geschriebenen Befund mit dem selben Informationsgehalt wie ein diktiertem Befund noch verlängern.

## 6 Zusammenfassung

Die Verwendung des digitalen Spracherkennungssystems SP 6000 als Unterstützung in der täglichen radiologischen Befundung konnte eingehend getestet werden. Schon nach dem Initialtraining konnte mit dem System bei mittleren Fehlerraten von 8,4 - 13,3 % effizient gearbeitet werden (Tabelle 4, siehe Seite 33). Grob geschätzt bedeutet dies ein falsch erkanntes Wort pro Satz. In der weiteren Entwicklung (nach der zweiten und dritten Adaptation) stellte sich eine durchschnittliche Erkennungsrate von ca. 98 % (entspricht einer Fehlerrate von 2 %) ein (Tabelle 7, siehe Seite 38). Die Menge der Diktate mit einer Erkennungsrate von 100 % lag nach der zweiten und dritten Adaptation zwischen 30 % und 40 % (Diagramm 9, Seite 38; Diagramm 11, Seite 39), d.h. 30 - 40 % der Diktate benötigten nur eine Kontrolle, jedoch keine Korrektur. Im Vergleich der Befundungszeit mittels der digitalen Spracherkennung mit dem konventionell geschriebenen Befund konnte abhängig von der individuellen Tastaturfertigkeit kein genereller Zeitvorteil des normalen schriftlichen Befundens ermittelt werden (Diagramm 16 und Diagramm 17, Seite 48). Für die meisten Radiologen sollte der Zeitgewinn jedoch bei der Nutzung des Spracherkennungssystems liegen. Außer dem individuellem Zeitvorteil besteht jedoch auch ein erhöhter Befundungskomfort, da der Blick nicht von den Aufnahmen abgewandt werden muß.

Nicht untersucht wurde die Möglichkeit, die Korrektur der erkannten Texte zentral von einer Sekretärin durchführen zu lassen. Insbesondere der Einfluß auf die Korrekturzeit und auf die Entwicklung der individuellen Fehlerrate wäre interessant.

Zur Entscheidung, auf welche Art ein Befund optimal erstellt werden kann, müssen Vor- und Nachteile gegeneinander abgewogen werden. Die Vor- und Nachteile des Schreibens und des konventionellen Diktates wurden bereits in der Einleitung erläutert. Nach unseren Erfahrungen hat das Spracherkennungssystem SP 6000 viele Vorteile:

- Der Radiologe hat die Hände frei (bei Nutzung eines Headphone-Sets) bzw. eine Hand bei Nutzung des Diktatmikrophons. Bei Nutzung des Mikrophons besteht keine Notwendigkeit den Monitor zu betrachten, da alle Funktionen durch das Mikrofon bedient und angezeigt (LED's) werden können.
- Im Vergleich zu der konventionellen Diktatkassette können im Diktat verschiedene Änderungen vorgenommen werden. Es gibt die Möglichkeiten des Einfügens, Übersprechens und Anhängens.
- Im Gegensatz zu anderen digitalen Spracherkennungssystemen (20, 31, 49) arbeitet das SP 6000 mit einer kontinuierlichen Spracherkennung. Das Diktieren kann in normaler

Sprachgeschwindigkeit ohne Pause zwischen den einzelnen Worten durchgeführt werden.

- Die Diktate werden kontinuierlich abgearbeitet. Es entsteht keine Pause zwischen der Beendigung des Diktates und dem Beginn der Spracherkennung, außer es werden gerade Diktate bearbeitet. Wenn das Diktat eine besondere Priorität zugewiesen bekommt, so wird es in der Liste der zu erkennenden Diktate an vorderste Stelle geschoben.
- Das System steht ständig zur Verfügung, exklusiv der Zeiten des Initialtrainings und der Adaptationen. Das Initialtraining bzw. die Adaptationen sollten daher abends gestartet werden.
- Es entstehen keine Tipp- oder Rechtschreibfehler.
- Die Diktat-PC's können sowohl für die Spracherkennung als auch für die konventionell geschriebene Befundung eingesetzt werden. Sollte der Spracherkennungsserver ausfallen, so besteht auf jedem Diktat-PC die Möglichkeit, das diktierete (und noch nicht erkannte) Diktat abzuhören und dabei zu schreiben bzw. schreiben zu lassen.

Es gibt natürlich auch Nachteile des Systems:

- Während das Diktieren mit dem SP 6000 ist nur an speziell eingerichteten Diktat-PC's möglich war, arbeitet die neueste Version dieses Systems hardwareunabhängig und kommt mit handelsüblichen Soundkarten zur D-A- bzw. A-D-Wandlung aus.
- Das Diktieren erfordert eine strenge Disziplin. Um ein korrekten Befund zu erhalten, muß man im Gegensatz zum konventionellen Diktat die komplette Interpunktion sowie eventuelle Textformatierungen mitdiktieren.
- Textänderungen im Diktat müssen während des Diktierens vorgenommen werden. Textänderungen nach Beendigung des Diktates sind erst nach der Übergabe des korrigierten Textes an das RIS- Befundsystem möglich.
- Neue Worte werden erst nach der nächsten Vokabelaktualisierung erkannt. Das bedeutet, daß diese Worte eventuell über mehrere Tage immer wieder korrigiert werden müssen, bis die nächste Vokabelaktualisierung durchgeführt wird.
- Das System arbeitet sprecherspezifisch. Jeder Nutzer muß das Initialtraining absolvieren.

Gerade Kollegen, die über längere Zeit Erfahrungen mit der konventionellen Diktiertechnik gemacht und ihren eigenen Diktierstil gefunden haben, werden sich mit den strengen Konventionen des SP 6000 schwer anfreunden können.

Nicht nur die Vor- und Nachteile einer Befundungstechnik bestimmen den Einsatz. Auch situative Gegebenheiten müssen bei dem Einsatz berücksichtigt werden.

So läßt sich das Diktieren (auch konventionell) trotz der maßgeblichen Vorteile nicht ausnahmslos favorisieren. In vielen Situationen ist eine Diktatbefundung durchaus vorstellbar, z.B. könnte der Arzt während der Durchführung der Ultraschalluntersuchung die Befundung vornehmen. Dieses Vorgehen hätte den Vorteil der sofortigen Befunderstellung. Da aber häufig Patienten zur Erstdiagnostik z.B. bei Tumorleiden sonographisch untersucht werden, könnten die Patienten durch die mündliche Befundung während der Untersuchung (auch bei einem nicht pathologischen Befund) stark verunsichert bzw. geängstigt werden. In unserem Institut ist von der arbeitsorganisatorischen Seite eine diktierende Befundung in der Ultraschalldiagnostik auch aus einem anderen Grunde nicht möglich. In dem Ultraschalluntersuchungsraum befinden sich zwei Ärzte. Während der eine Arzt die aktuelle Untersuchung vornimmt, befundet der zweite Arzt die vorhergehende Untersuchung, so daß im Wechsel befundet und untersucht wird. Außer dem störenden Einfluss auf die laufende Untersuchung wäre der Datenschutz nicht gewährleistet, da der Patient den Befund der vorherigen Untersuchung mit anhören könnte.

Unter Betrachtung der ermittelten Ergebnisse kann das digitale Spracherkennungssystem SP 6000 als vorteilhafte Alternative zur schnellen Erstellung radiologischer Befunde angesehen werden. Durch die Hybridfunktion des Diktat-PC's, der als konventionelle Diktierstation, zur Spracherkennung und zum normalen Schreiben genutzt werden kann, bietet sich das SP 6000 für jede Art der Befunderstellung an.

Aus unserer Sicht gibt es verschiedene Möglichkeiten der Verbesserung des Systems. Günstig wäre eine systemeigene Überwachung der angefallenen Diktatzeiten der einzelnen Nutzer, so daß bei Erreichen einer einstellbaren Zeit das akustische Training automatisch gestartet werden kann. Ähnlich könnte mit der Vokabelaktualisierung verfahren werden; nachdem eine bestimmte Anzahl neuer Worte erfaßt worden ist, könnte die Vokabelaktualisierung durchgeführt werden. Im Rahmen der Weiterentwicklung des Systems wurde bei Philips ein mobiles digitales Diktiersystems entwickelt. Mit einem derartigen Diktiersystem hat man die Möglichkeit, mobil erstellte Diktate an den Diktat-PC's direkt (über eine digitale Schnittstelle) in das Spracherkennungssystem zu übertragen.

Der Optimismus, der in der Literatur (11, 49) im Zusammenhang mit der Entwicklung und dem Einsatz der digitalen Sprachverarbeitung in der Medizin zu finden ist, wird von uns geteilt. Wenn man davon ausgeht, daß die stürmische Entwicklung der Mikroelektronik, Prozessortechnik und auch der künstlichen Intelligenz weiter anhält, so ist durchaus vor-

stellbar, daß auch bei komplexen Spracherkennungssystemen wie dem SP 6000 die Hürde der Sprecherabhängigkeit überwunden werden wird.

## 7 Literatur

- 1 Atal, B.S.: Speech technology in 2001: New research directions in Proc-Natl-Acad-Sci-U-S-A 92, 22: 10046- 10051, 1995
- 2 Bates, M.: Models of natural language understanding in Proc-Natl-Acad-Sci-U-S-A 92, 22: 9977- 9982, 1995
- 3 Bergeron, B.; Locke, S.: Speech Recognition as a User Interface in M.D.Computing 7,5: 329- 334, 1990
- 4 Cohen, P.R; Oviatt, S.L. : The role of voice input for human-machine communication in Proc-Natl-Acad-Sci-U-S-A 92, 22: 9921- 9927, 1995
- 5 De Bruijn, L.M. et al: Speech interfacing for Diagnosis reporting systems: an overview in Computer-Methods-and-Programs-in-Biomedicine 48,1-2: 151-156, 1995
- 6 Deller, J.R et al: Diskrete Time-Processing of Speech Signals , Macmillan, New York, 1993
- 7 Detmer, W.M. et al : A Continuos-speech Interface to a Decision Support System: An Evaluation Using a Wizard-of-Oz Experimental Paradigm in J-Am—Med-Inform-Assoc 2, 1: 46- 57, 1994
- 8 Dugast, C. : Large-vocabulary recognition in Philips Journal of Research 49,4: 353- 366, 1995
- 9 Fant, G.: Acoustic Theorie of Speech Production, Mouton, The Hague, The Netherlands, 1960
- 10 Flanagan, J.L.: Speech analysis, synthesis, and perception, Springer, New York, 1972
- 11 Flanagan, J.L.: Research in speech communication in Proc-Natl-Acad-Sci-U-S-A 92, 22: 9938- 9945, 1995
- 12 Furui, S.: Toward the ultimative synthesis/recognition system in Proc-Natl-Acad-Sci-U-S-A 92, 22: 10040- 10045, 1995
- 13 Gold, B.; Rader,C.M.: Digital Processing of Signals, McGraw-Hill, New York, 1969

- 14 Haeb- Umbach, R. et al : Speech recognition algorithms for voice control interfaces in Philips Journal of Research 49,4: 381-397, 1995
- 15 Hansen, G.C. et al : Voice recognition System (Letter) in Radiology 169(2) : 580, 1988
- 16 Hirschmann, L.: The role of natural language understanding in Proc-Natl-Acad-Sci-U-S-A 92, 22: 9970- 9976, 1995
- 17 Jelinek, F.: Training and search methods for speech recognition in Proc-Natl-Acad-Sci-U-S-A 92, 22: 9964- 9969, 1995
- 18 Kamm, C. User interfaces for voice applications in Proc-Natl-Acad-Sci-U-S-A 92, 22: 10031- 10037, 1995
- 19 Kato, Y.: The future of voice-processing in the world of computers and communications in Proc-Natl-Acad-Sci-U-S-A 92, 22: 10060- 10063, 1995
- 20 Kolles, H.; Feiden, W.: Computergestützte Spracherkennung in der diagnostischen Pathologie in Pathologie 16: 439- 442, 1995
- 21 Landau, J.A. et al Automatic Speech Recognition - can it improve the man- machine interface in medical expert systems? In Int-J-Biomed-Comput 24,2: 111-117, 1989
- 22 Levinson, S.E.: Speech recognition technology: a critique in Proc-Natl-Acad-Sci-U-S-A 92, 22: 9953- 9955, 1995
- 23 Levinson, S.E.; Fallside, F.: Speech technology in the year 2001 in Proc-Natl-Acad-Sci-U-S-A 92, 22: 10038- 10039, 1995
- 24 Levitt, H.: Processing of speech signals for physical and sensory disabilities in Proc-Natl-Acad-Sci-U-S-A 92, 22: 9999- 10006, 1995
- 25 Makhoul, J.; Schwartz, R.: State of the art in continuous speech recognition in Proc-Natl-Acad-Sci-U-S-A 92, 22: 9956- 9963, 1995
- 26 Maragos,P.A. : Proceedings of the IEEE International Conferenze on Acoustics, Speech and Signal, Toronto, 417-420, 1991
- 27 Matumoto, T. et al: Automatic Radiologic Reporting System using Speech Recognition in Med-Prog-Technol 12, 3-4: 243- 257, 1987

- 28 Marcus, M.: New trends in natural language processing: Statistical natural language processing in Proc-Natl-Acad-Sci-U-S-A 92, 22: 10052- 10059, 1995
- 29 Meijer, G.A.; Baak, J.P.: Reporting by Digital Speech Recognition in Human Pathology 26: 813- 815, 1995
- 30 Moore, R.C.: Integration of Speech with natural language understanding in Proc-Natl-Acad-Sci-U-S-A 92, 22: 9983- 9988, 1995
- 31 Mrosek, B. et al : Computergestützte Spracherkennung und Ausdruck von Röntgenbefunden in Fortschr Röntgenstr. 159,5: 481- 483, 1993
- 32 Murchie, C. J.; Kenny, G.N.C.: Comparisation of Keyboard, Light pen and Voice Recognition as methods of Data Input in Int-J-Clin-Monit-Comput 5, 4: 243- 246, 1988
- 33 Mutschler, H., Schöller, B.: Spracheingabe für Behinderte in Rehabilitation 25, 3: 123- 127, 1986
- 34 Nakatsu, R.; Suzuki, Y.: What does voice-processing technology supports today? in Proc-Natl-Acad-Sci-U-S-A 92, 22: 10023- 10030, 1995
- 35 Oberteuffer, J.A.: Commercial applications of speech interface technology: An industry at the treshhold in Proc-Natl-Acad-Sci-U-S-A 92, 22: 10007- 10010, 1995
- 36 O'Hara, S.P. et al: Speech Recognition and the Clinical Microbiology Laboratory, Med-Lab-Sci 49: 20-26, 1992
- 37 Oppenheim, A.V.; Schafer, R.W.: Digital Signal Processing, Prentice-Hall, Englewood Cliffs, NJ, 1975
- 38 Pierce, J.R.: in J.Acoust.Soc.Am. 47, 1049- 1050, 1969
- 39 Rabiner, L.R.; Schafer, R.W.: Digital Processing of Speech Signals, Prentice-Hall, Englewood Cliffs, NJ, 1978
- 40 Rabiner, L.R.; Juang, B.H.: Fundamentals of Speech Recognition, Prentice-Hall, Englewood Cliffs, NJ, 1993
- 41 Rabiner, R.L. : Voice communication between humans an machines - an introduction in Proc-Natl-Acad-Sci-U-S-A 92, 22: 9911- 9913, 1995

- 
- 42 Roe, D.B. Deployment of human-machine dialogue systems in Proc-Natl-Acad-Sci-U-S-A 92, 22: 10017- 10022, 1995
  - 43 Sarnat, A. J. : Computerized Speech Recognition for anaesthesia recordkeeping in Med-Instrum 17, 1: 25-27, 1983
  - 44 Schafer, R.W. : Scientific bases of human-machine communication by voice in Proc-Natl-Acad-Sci-U-S-A 92, 22: 9914- 9920, 1995
  - 45 Seelbach, C.: A perspective on early commercial applications of voice-processing technology for telecommunications and aids for the handicapped in Proc-Natl-Acad-Sci-U-S-A 92, 22: 9989- 9990, 1995
  - 46 Shillitoe, E.J.: A DNA Sequence Editor with Voice Input and Output in Biotechniques 17: 1144- 1146, 1994
  - 47 Sluijter, R.J. et al : State of the art and trends in Speech coding in Philips Journal of Research 49,4: 455-488, 1995
  - 48 Spacone, A. B.: Microcomputer Voice- recognition Program in a Hospital Emergency Department in J-Soc-Health-Syst 1, 2: 111- 118, 1989
  - 49 Teplitz, C. et al: Automated Speech Recognition Anatomic Pathology (ASAP) Reporting in Seminars in Diagnostik Pathology 11, 4: 245 - 252, 1994
  - 50 Weinstein, C.J.: Military and government applications of human-machine communication by voice in Proc-Natl-Acad-Sci-U-S-A 92, 22: 10011- 10016, 1995
  - 51 Wilpon, J.G.: Voice-processing technologies-Their application in telecommunications in Proc-Natl-Acad-Sci-U-S-A 92, 22: 9991- 9998, 1995
  - 52 Zier, M.A.: More on Speech Recognition Systems in M.D. Computing 8,2: 72, 1991

## **Eidesstattliche Erklärung**

Hiermit erkläre ich, Holger Arndt, geb. am 16.6.1967, an Eides statt, die vorliegende Arbeit selbständig ohne Mitarbeit Dritter angefertigt zu haben.

Berlin, den 24.5.98

## **Danksagung**

Mein besonderer Dank gilt Herrn Prof. Dr. med. B. Hamm, Direktor des Institutes für Röntgendiagnostik der Charité Berlin für die Überlassung des Themas.

Besonders möchte ich außerdem Herrn Dr. med. Dipl. phys. J. Petersein danken, da er einen großen Anteil an der Datenerfassung trug und mir immer mit Rat und Tat hilfreich zur Seite stand.

Desweiteren danke ich Herrn Dr. med. D. Stockheim für die Unterstützung bei der Datenerfassung. Herrn Dipl. Ing. P. Gregor und Herrn Stark von PMS möchte ich für die Hilfe bei der Betreuung und Instandhaltung des Systems danken.

## Lebenslauf

Name, Vorname: Arndt, Holger  
Geburtsdatum: 16.6.1967  
Geburtsort: Berlin

Eltern: Dr. rer. nat. Arndt, Alfred, geb. am 12.3.1941, Wissenschaftler  
Dipl. med. Arndt, Karin, geb. Michalke, geb. am 25.4.1943,  
gest. am 23.2.1992, Fachärztin für Kinderheilkunde

Wohnung: Am Schloßpark 10, 13187 Berlin  
Familienstand: verheiratet, 2 Kinder

Schulbildung: 1974 - 1984 15. Polytechnische Oberschule Berlin  
Prädikat: „mit Auszeichnung“  
1984 - 1986 Erweiterte Oberschule „Carl von Ossietzky“  
1986 Abitur mit dem Prädikat: „sehr gut“

Hochschulstudium: 1989 - 1995 Medizinstudium an der Humboldt-Universität zu  
Berlin, medizinische Fakultät (Charité)  
Oktober 1996 Abschluß mit dem Prädikat: „Gut“