

6 Algorithmen bei der multimedialen Erkundung von Grafiken

Landläufig wird davon ausgegangen, daß Grafiken für sich selbst sprechen und man zu ihrer Interpretation keine weiteren Hilfsmittel brauche. Daß diese Annahme sogar für gute Bilder nicht grundsätzlich zutrifft, hat z.B. Weidenmann bewiesen [WEIDENMANN 89]: Fehlt eine zusätzliche semantische Untermauerung der Grafik oder ist sie ungeeignet, so sind auch gut gestaltete Grafiken nicht oder nicht richtig zu interpretieren. Die angesprochenen Untersuchungen beziehen sich auf visuelle Grafiken, für die es Qualitätskriterien gibt und deren Wahrnehmung schon recht gut erforscht ist.

Für haptische Grafiken, insbesondere solche räumlicher Gegenstände, gibt es bisher weder diese noch jene: Die rudimentär vorhandenen Qualitätskriterien [EDMAN 92] beziehen sich meist nur auf lexikalische Eigenschaften wie Linienabstand usw.. Da über die Wahrnehmungsverfahren nicht-visueller Grafiken auch noch recht wenig bekannt ist und es auf diesem Gebiet kaum eine kulturelle Tradition gibt, ist zumindest nicht davon auszugehen, daß nicht-visuelle Grafiken leichter zu interpretieren sind als visuelle.

Trotzdem muß es das Ziel des Entwurfs haptischer Grafiken sein, leicht erkennbare, informative und dem jeweiligen Zweck angemessene Grafiken zu erstellen. Eine solche Grafik kann weitestgehend autark erfaßt werden. Liegt sie z.B. auf Schwellpapier vor, so kann der Erfasser im Idealfall alle dargestellten Gegenstände, ihre Struktur und räumliche Anordnung ohne weitere Hilfsmittel erkennen. *Nicht* erkennen kann er auch in diesem Fall grafisch nicht darstellbare Eigenschaften wie z.B. die Materialbeschaffenheit.

Um sowohl die Erkennung der dargestellten Gegenstände als auch die der nicht darstellbaren Eigenschaften zu erleichtern bzw. zu ermöglichen, bieten sich textuelle Zusatzinformationen an, die mit den grafischen Elementen verknüpft sind. Während solche Zusatzinformationen in visuellen Grafiken mitunter in Form von Beschriftungen integriert sein können, haben sich Beschriftungen (in Braille-Schrift) in haptischen Grafiken nicht bewährt, da sie zu viel Platz beanspruchen und/oder andere grafische Elemente stören würden [KORDON 95].

In diesem Kapitel werden Verfahren zur Unterstützung der Erkundung von haptischen Grafiken dargestellt, die entweder schon beim Rendern oder erst bei der eigentlichen Erkundung zum Einsatz kommen.

6.1 Selbständige Erkundung

Wegen der Heterogenität der Gruppe der Blinden und der nur vereinzelt praktizierten Ausbildung an taktilen Grafiken lassen sich kaum allgemeingültige Aussagen darüber machen, wie Blinde typischerweise bei der Erkundung von Grafiken vorgehen. Im folgenden wird ein prototypisches Vorgehen skizziert, das die einzelnen Verfahren beinhaltet, auf die dann im Anschluß einzugehen sein wird.

Der (prototypische) blinde Erfasser ertastet taktile Grafiken zunächst mit der ganzen Handfläche und untersucht dann die Details. In welcher Weise dabei vorgegangen wird, läßt sich schwer vorhersagen. Es können zunächst die beim Überblicksertasten als interessant empfundenen Bereiche abgetastet werden, oder die Zeichnung kann systematisch von links oben nach rechts unten (oder von innen nach außen oder ganz anders) abgetastet werden.

Diesem Verfahren ist gemeinsam, daß sich die detaillierte Erkundung nur mit recht kleinen Flächen beschäftigen kann. Es wird mit einem bis zwei Fingern eine Linie oder eine Fläche

untersucht, wobei gegebenenfalls ein weiterer Finger oder die andere Hand zur Orientierung an einer Stelle bleibt. Trotz dieser Orientierungsunterstützung ergeben sich oft Probleme, eine bestimmte Stelle einer noch nicht erkannten Grafik wiederzufinden.

Sobald die Grafik erkannt, also die in ihr dargestellte Szene erfaßt wurde, ist auch das Wiederfinden schon ertasteter Objekte leicht möglich. Dies hängt offensichtlich mit dem dann vorhandenen korrekten mentalen Modell des Dargestellten zusammen. Zur Konstruktion des mentalen Modells ist andererseits wieder die Erkennung von Gegenständen notwendig, damit diese einen Kontext für nachfolgend erfaßte weitere Objekte bilden können

Inwieweit die dargestellten Gegenstände erkannt werden, hängt von der Erfahrung des Blinden, dem Kontext und (ganz entscheidend) von der Qualität der Zeichnung ab.

Um die Erkennung zu vereinfachen und zu beschleunigen, haben sich Ansätze wie Nomad [PARKES 88] und AudioTouch [LÖTZSCH 94] oder auch TGuide [FISCHER 96] durchgesetzt. Hierbei wird die zeichnerisch dargestellte Information mit einer Datenbank hinterlegt, die positionsbezogen über ein Touchpad oder ähnliche Geräte abgefragt werden kann.

Die Ausgabe der hinterlegten Information erfolgt in der Regel über einen Sprach-Synthesizer. Auf die Mechanismen in diesem Zusammenhang wird noch detaillierter einzugehen sein (Siehe Kapitel 7.4). Grundsätzlich kann der Benutzer auf diese Information zugreifen, indem er an einer beliebigen ihn interessierenden Stelle auf die Grafik (bzw. ein darunter befindliches Tasttablett (Touchpad)) drückt oder ein Hilfsmittel wie den Digitalisierstift zur Markierung verwendet. Das verwendete Gerät liefert dem Erkundungs-Hilfsprogramm die zweidimensionalen Koordinaten dieser Stelle.

Zunächst ist der Aufbau der Datenbasis, also die Gewinnung der Daten, näher zu untersuchen und sind die gewählten Datenstrukturen einer kritischen Bewertung zu unterziehen.

Die hinterlegte Datenbasis wird (wie auch praktisch alle taktilen Zeichnungen) bisher noch komplett von Hand erzeugt, d.h. die Regionen der Zeichnung, die einen bestimmten Gegenstand zeigen, werden manuell bestimmt (z.B. indem sie per Digitalisierstift und -stift umrandet werden). Auch die den einzelnen Bereichen (Regionen) hinterlegte Information, in der Regel die Bezeichnung des Gegenstands, wird manuell eingegeben und mit der entsprechenden Region verknüpft.

Während die textuelle Information meist in Form einer einfachen Textdatei vorliegt, die in jeder Zeile genau einen (textuellen) Bezeichner enthält, bieten sich für die Grafik zunächst zwei Datenstrukturen an:

- Die Grafik kann als *Rastergrafik* repräsentiert sein, wobei jedes Pixel die Information darüber trägt, welcher Text mit ihr assoziiert ist (vgl. Abbildung 6-1). Der Vorteil dieser Repräsentation ist der nunmehr sehr schnell mögliche Zugriff auf den Text, da die Textposition sich direkt aus der Grafik entnehmen läßt. Da die Rastergrafik allerdings in einer recht hohen Auflösung vorliegen muß, um die feinen Strukturen typischer Liniengrafiken verwalten zu können, ist die sich ergebende Datenmenge sehr hoch²⁸. Die Struktur einer einfachen zweidimensionalen Matrix bietet außerdem keinerlei Informationen darüber, zu welcher Art von grafischen Objekten der untersuchte Punkt gehört (Linie, Kreis, Polygon, ...). Auch weitere Daten über das Objekt, z.B. seine räumliche Position vor oder hinter anderen dargestellten Gegenständen, lassen sich aus der Pixelrepräsentation nicht mehr gewinnen. Diese Datenstruktur entspricht weitgehend dem visuellen Bild auf einem Bildschirm, bei dem ebenfalls von jedem Pixel nur seine Farbe

²⁸ Legt man eine Graphik in der Größe eines DIN A4 Blattes und eine Pixelgröße von 1 mm² zugrunde, so ergibt sich, daß 62.370 Pixel verwaltet werden müssen. In der Praxis kann diese Auflösung allerdings sogar noch zu niedrig sein, so daß sich noch größere Datenmengen ergeben.

bekannt ist. In AudioTouch wird eine solche Rastergrafik als zweidimensionaler Indexraum verwendet, um die zugehörigen Textinformationen identifizieren zu können.

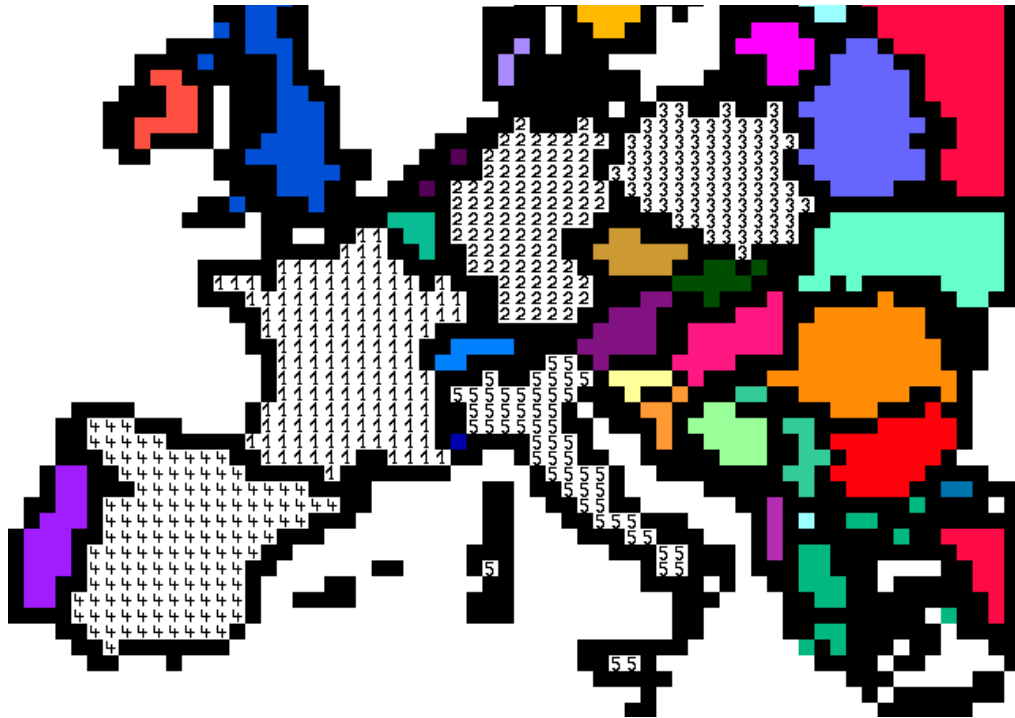


Abbildung 6-1 Europa als (grobe) Rastergrafik. Beispielhaft ist für fünf Länder gezeigt, wie hinter jedem Bildpunkt ein Index auf eine Datenbank abgelegt ist, die etwa folgende Form hätte:

- 1 Frankreich
- 2 Deutschland
- 3 Polen
- 4 Spanien
- 5 Italien

...

Die Texte zu den Indizes können per Sprachsynthese ausgegeben werden.

- Die Grafik kann auch als *Vektorgrafik* repräsentiert sein. In diesem Fall werden für die dargestellten Gegenstände, die (zweidimensionalen) geometrischen Primitive (Linien, Kreise, Polygone, ...), aus denen die Darstellung besteht, mit Position, Größe und Ausrichtung gespeichert (vgl. Abbildung 6-2). Zu jedem geometrischen Grundobjekt muß ein Verweis auf die textuelle Beschreibung existieren. Da aus der Koordinaten-Information, die die Benutzerinteraktion liefert, nicht unmittelbar auf ein geometrisches Objekt der Zeichnung geschlossen werden kann (im Gegensatz zur Rastergrafik, bei der (in der Realisierung in AudioTouch) jedem Pixel eine Verknüpfung auf das Objekt hinterlegt ist, das es darstellt), muß zunächst ermittelt werden, welches Objekt der Zeichnung sich an der übermittelten Stelle befindet (Details unter 6.3). Dies ist unter Umständen ein aufwendiger Prozeß. Auf der anderen Seite ist der Speicherbedarf von Vektorgrafiken meist geringer als der vergleichbarer Rastergrafiken.

Hinzu kommt, daß sich Vektorgrafiken verlustfrei beliebig skalieren lassen und daher auch die Zuordnung einer interessierenden Stelle zu (beispielsweise) einer Linie flexibel gehandhabt werden kann: Ist die Positionierung genau genug, kann die Toleranz bei der Objektidentifizierung herabgesetzt und so der Zugang zu feineren Details ermöglicht werden. Schließlich gestattet die Repräsentation als Vektorgrafik auch eine (implizite) Schichtung der Linien „von hinten nach vorn“, so daß diese Tiefeninformation prinzipiell auch aus diesen Daten gezogen werden kann. TRender erzeugt solche Vektorreprä-

sentationen, die dann von TGuide direkt verwendet oder (nach einer Umwandlung in das dort verwendete Rasterformat) mit AudioTouch präsentiert werden können.

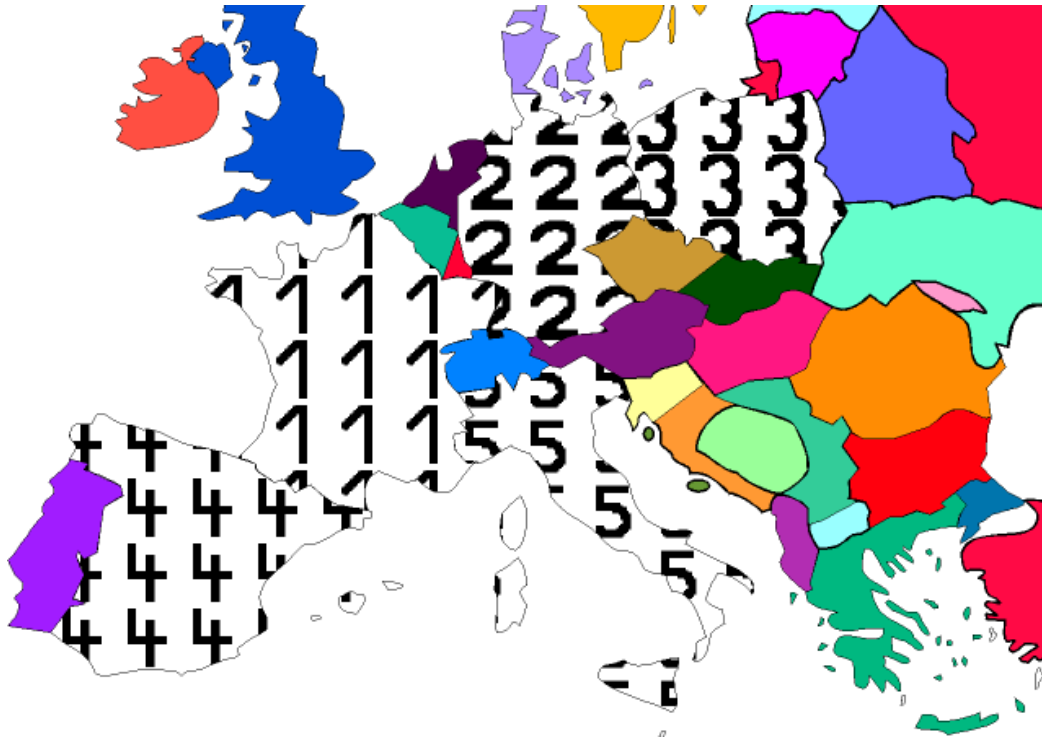


Abbildung 6-2 Europa als Vektorgrafik. Es wird zunächst festgestellt, innerhalb welcher Flächen-Grenzen Informationen abgefragt wurde. Aus dem Index dieser Fläche wird in der Datenbank aus Abbildung 6-1 der auszugebende Text abgeleitet und vorgelesen.

Beim Erkunden der haptischen Grafik, die auf einem geeigneten Tablett (Tasttablett oder Digitalisiertablett bzw. Nomad) liegt, steht dem Blinden nun neben der grafischen Information und den darin kodierten räumlichen Eigenschaften auch semantische Information zur Verfügung, insbesondere Objektbezeichnungen. Gerade durch die semantische Anreicherung und Hinterlegung wird die Bildung und Aktualisierung des mentalen Modells unterstützt, wodurch der Kontext wahrgenommen wird, der bei der weiteren Erkundung nützlich sein kann.

In diesem Zusammenhang ist es bedeutsam, sich zu erneut vergegenwärtigen, daß die Information, an einer Stelle befinde sich „ein Tisch“, die auch rein textuell vermittelt werden könnte, nicht von der gleichen Qualität ist, wie die Kombination aus der (textuellen) Information „dies ist ein Tisch“ mit der grafisch dargestellten Repräsentation des gemeinsamen Tisches in all seinen geometrischen Details.

Wenn auch die Erkundung der haptischen Grafik mit Unterstützung durch die verbale Ausgabe hinterlegter Information vereinfacht wird, so ist doch nicht ausgeschlossen, daß wichtige Details verborgen bleiben oder daß die Erkundung unnötig erschwert wird, weil die Objekte beim Abtasten in einer ungeeigneten Reihenfolge erfaßt werden. Im folgenden Abschnitt wird auf diese Problematik eingegangen und ein Verfahren gezeigt, mit dem die Erkundung durch den Blinden aktiv unterstützt wird.

6.2 Geleitete Erkundung

Sehende können mit mehr oder weniger subtilen Verfahren dazu gebracht werden, die Elemente einer Grafik in einer bestimmten Reihenfolge zu erkunden. Insbesondere kann

die Aufmerksamkeit auf bestimmte Details gelenkt werden. Dies kann im einfachsten Fall ein Pfeil oder eine andere Markierung sein, die sich vom Rest der Zeichnung abhebt.

Da eine Formenerkennung (Pfeil o.ä.) bei der taktilen Erkundung stets sequentiell ablaufen muß, also eine Form nicht auf Anhieb wie beim Sehen erkannt werden kann, nutzen solche Markierungen nichts für eine Beeinflussung des Erkundungsvorgangs. Eine geleitete Erkundung ist auf Hinweisformen angewiesen, die nicht in der Zeichnung selbst verankert sind.

Das Problem kann darauf reduziert werden, eine Möglichkeit zu finden, die Hand des Blinden zu einer bestimmten Stelle der Zeichnung zu führen. Dabei kann man zwei Klassen von Verfahren unterscheiden:

- Die Hand des Benutzers kann durch mechanische Kraftausübung bewegt werden. Die Hand wird also passiv bewegt.
- Dem Benutzer kann eine Richtung angegeben werden, in die er seine Hand dann selbst (aktiv) bewegt.

Für das erste Verfahren würde sich beispielsweise ein Roboter-Arm oder ein Pantograph [RAMSTEIN & HAYWARD 94] eignen, wobei die Hand oder zumindest ein Finger in einer Halterung am Ende des Arms befestigt werden müßte. Die Akzeptanz einer solchen extern induzierten Bewegung ist allerdings recht gering²⁹. Hinzu kommt noch, daß eine solche Führung eher einem passiven als einem aktiven Tasten entspricht. Die Probleme der passiven Tastens wurden in Abschnitt 2.3 behandelt und träten bei der skizzierten Vorgehensweise analog auf.

Beim zweiten Verfahren, nämlich dem Blinden eine Richtung *vorzuschlagen*, in die er seine Hand bewegen sollte, lassen sich verschiedene Techniken einsetzen:

Verbale Richtungshinweise: Dem Blinden wird die Richtung, in die er seine Hand bewegen soll sprachlich übermittelt ("weiter rechts, weiter nach links oben, ..."). Die Richtungen könnten auch als Himmelsrichtungen angegeben werden („Ost, Nordwest, ...“). Beide Varianten haben Probleme: Angaben wie „rechts“ oder „links“ beziehen sich immer auf eine Grundrichtung; diese kann die Achse des Papiers sein oder auch die momentane Bewegungsrichtung der Hand. Dieser Zusammenhang wird in Abbildung 6-1 deutlich. Die Umsetzung von Himmelsrichtungen kennt dieses Problem nicht, da sich Himmelsrichtungsangaben (bei Karten) immer auf die Achse des Blattes beziehen; allerdings ist sie den meisten Menschen nicht so vertraut wie das Rechts-Links-Schema und verlangt daher größere kognitive Arbeit. Verbale Richtungshinweise jeder Art verlangen, zunächst die Richtungsangaben kognitiv umzusetzen und dann daraus Handbewegungen zu planen und durchzuführen. Eine unmittelbare Ohr-Hand-Koordination findet nicht statt.

²⁹ Verschiedene Gespräche mit Blinden anlässlich der Evaluation von TGuide (Abschnitt 8.4) und mit Fachleuten des Gebiets stützen diese Aussage.

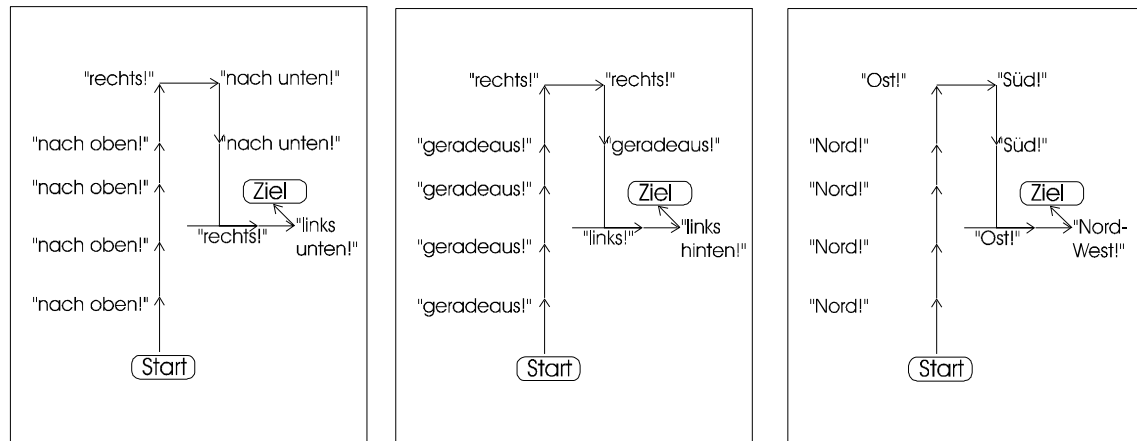


Abbildung 6-3 3 Varianten, eine Erkundung entlang einer Linie verbal zu unterstützen:
 links: absolute Richtungshinweise (Annahme: Zeichnung steht senkrecht)
 Mitte: relative Richtungshinweise („Straßenkartennavigation/Beifahrer-Unterstützung“)
 rechts: absolute Navigation mit Himmelsrichtungen (Annahme: Zeichnung liegt waagrecht)

Nonverbale akustische Signale: Durch Variation der Tonhöhe und einer Stakkato-Frequenz können ebenfalls vergleichbare Hinweise gegeben werden. Auch hier ist ein Lernaufwand nötig, und die kognitive Ebene läßt sich nicht umgehen.

Beide hier genannten Techniken belegen den akustischen Wahrnehmungskanal und verhindern oder erschweren den Einsatz der Hilfsmittel für die selbständige Erkundung wie er in Abschnitt 6.1 beschrieben ist. Die folgenden Techniken verwenden nicht den akustischen sondern den haptischen Sinn zur Richtungsangabe.

Aktiv-kinästhetische Führung: Wenn man die oben erwähnte Methode der induzierten Bewegung variiert, könnte man mit einem Plotter, Roboter-Arm oder einem Pantograph-Gerät der Hand eine bevorzugte Bewegungsrichtung nahelegen, ohne dabei die Hand selbst zu bewegen. Wenn die angewendete Kraft allerdings zu groß gewählt wird, ergeben sich die genannten Akzeptanzprobleme bei den Benutzern, da nicht Vorschläge unterbreitet werden, sondern der Benutzer zum passiven Hinnehmen/Ertragen einer Bewegung genötigt wird, die er noch nicht einmal selber ausführen muß. Zusätzlich sind die technischen Probleme recht erheblich. Eine solche Kraftausübung zur Navigationsunterstützung verbaut außerdem die Möglichkeit, die Orientierung von Flächen und Kanten im Raum durch Kräfte entsprechend der Schwerkraft darzustellen (siehe Abschnitte 5.3 und 9.2).

Aktiv-taktile Führung: Ein einer Maus ähnliches Gerät, das der Blinde unter der Handfläche über das Papier führt, kann ihm eine Richtung (von z.B. acht möglichen) vorschlagen, indem es an der entsprechenden Stelle einen Vibrator-Stift aktiviert. Die Hand kann in Richtung der Vibration bewegt werden, bis sie über der Zielposition liegt. Dabei behält der Blinde stets die Kontrolle über seine Hand. Er muß dem Vorschlag nicht folgen. Ein entsprechendes Werkzeug wird im Abschnitt 7.4 beschrieben.

Neben den Verfahren, die das Navigieren durch eine Richtungsangabe, also durch relative, von der Position der Hand abhängige Signale, unterstützen, sind auch solche möglich, die absolut arbeiten, also unabhängig von der Handposition immer die Zielposition direkt zeigen. Dies kann z.B. akustisch geschehen:

Raum- bzw. flächenakustische Positionsangabe: Durch Geräuschquellen, die unter der Zeichnung angebracht werden, könnten dem Blinden unmittelbar Stellen auf der Zeichnung akustisch gezeigt werden. Er müßte seine Hand lediglich über die Geräuschquelle bewegen. Hier ergeben sich zunächst technische Probleme, die jedoch nicht unlösbar erscheinen. Die

Ohr-Hand-Koordination könnte unmittelbar über die Positionswahrnehmung des Ohres geschehen. Das Verfahren setzt voraus, daß das menschliche Richtungshören hinreichend genau arbeitet. Untersuchungen über die Fähigkeit des Menschen, eine Geräuschquelle zu orten, lassen den Schluß zu, daß ein solches Verfahren zur Orientierungsunterstützung geeignet ist, wenn nicht zu viele Geräuschquellen gleichzeitig aktiv sind [CRISPIEN & PETRIE 93].

Alle Verfahren zur geleiteten Erkundung gehen davon aus, daß eine Instanz außerhalb der Grafik selbst Ziele vorgibt, auf die sie den Benutzer hinführt. Welche Verfahren zur Auswahl dieser Ziele angewendet werden, hängt vom Kontext der Grafik und ihrem Inhalt sowie letztendlich von der ihr unterlegten Information ab. Im Abschnitt 7.4 wird ein entsprechendes System kurz beschrieben.

6.3 Erkunden von Bildern mittels Sprache

In der Mensch-Computer-Interaktion kommt die Sprache vor allem in Form von synthetischer Sprache als *Ausgabemedium* vor. Insbesondere für Blinde spielt die Sprachausgabe häufig die Rolle des primären Ausgabemediums für textuelle Informationen [FELLBAUM 84, FELLBAUM ET AL. 92]. Seit die automatische Spracherkennung (zumindest sprecherabhängig) zufriedenstellend funktioniert, kann gesprochene Sprache auch als *Eingabemedium* verwendet werden.

Der Ansatz, Blinden die für Sehende visuell dargestellte Information sprachlich zu vermitteln, liegt nahe. Was für gedruckte Texte angemessen sein mag, kann jedoch nicht automatisch für grafische Information übernommen werden (siehe auch [FELLBAUM ET AL. 92, WETTER 95]).

Die Grafik als Medium unterscheidet sich, wie in den ersten Kapiteln dieser Arbeit gezeigt, grundlegend von der linearen Form eines Textes. Daraus ergibt sich, daß ohne einen Medienwechsel und mithin eine Änderung der Darstellungsklasse ein Zugang zu Grafiken für Blinde per Sprachausgabe allein nicht möglich ist. Hinzu kommt, daß man aus einer formalen Beschreibung von geometrischen Eigenschaften (noch) keine natürlichsprachliche Darstellung generieren kann. Dies trifft insbesondere dann zu, wenn das geometrische Modell keine oder ungeeignete semantische Informationen enthält: Ein Tischbein, das nur als `box` oder gar als `object4711` bezeichnet wird, kann nicht als „Tischbein“ sprachlich ausgegeben werden.



Abbildung 6-4 Einige Probleme bei der Bildbeschreibung: Eine einfache Szene wie diese ist nicht leicht eindeutig zu beschreiben: Dies wird anhand einiger Fragen deutlich, die an das System gestellt werden könnten: Steht der Baum *vor* oder *links* vom Auto? Wenn der Hund hinter dem Auto steht und es anschaut, schaut er dann nach vorn? Wenn der Fotograf hinter dem Auto steht und es fotografiert, ist dann die rechte Tür des Autos auf dem Foto sichtbar? Wenn der Vogel in die entgegengesetzte Richtung läuft wie das Auto, läuft er dann auf das Auto zu? Oder von ihm weg?

In dieser Arbeit wird auch der Einsatz von Sprache als wichtiges unterstützendes Medium bei der Erkundung von Zeichnungen untersucht. Sprache kann hierbei sowohl als Eingabe- als auch als Ausgabemedium verwendet werden (siehe auch [FELLBAUM 96, FELLBAUM 97]).

Sprachausgabe ist in folgenden Situationen anwendbar:

- Die *Ausgabe von Objektbezeichnungen*, die den zweidimensionalen grafischen Elementen hinterlegt sind, ist der klassische Fall der Sprachunterstützung bei der selbständigen Erkundung [PARKES 88, LÖTZSCH 94]. Die auszugebenden sprachlichen Elemente (Wörter, Sätze) liegen in der der Grafik hinterlegten Datenbasis vor und können direkt per Sprachsynthese ausgegeben werden. Die Qualität der Sprachsynthesysteme wird von der überwiegenden Mehrheit der Benutzer akzeptiert. Falls die Datenbasis manuell erstellt wird (wie z.B. bei AudioTouch [AUDIO TOUCH 95]), kann die Objektbezeichnung auch digital aufgenommen („gesampelt“) und dann bei der Erkundung per Sprachwiedergabe in naturgetreuer Qualität ausgegeben werden. Dem Vorteil der höheren Qualität steht der Nachteil der geringeren Flexibilität und des höheren Datenvolumens gegenüber. Wirtschaftlich schrumpft der Vorsprung der Ton-(Sprach-)Wiedergabe gegenüber der Sprachsynthese zusehends, da letztere immer preiswerter wird und oft sogar zum Lieferumfang eines Rechners gehört.
- Die *Ausgabe ergänzender Textinformation*, die aus der als Referenz verwendbaren Objektbezeichnung gewonnen werden kann, wenn die Datenbasis entsprechend strukturiert und inhaltlich aufgebaut ist, kann auf dem gleichen Weg erfolgen. Hier bietet sich allerdings die Sprachwiedergabe weniger an, weil sie voraussetzt, daß der gesamte Text zunächst gesampelt wird. Voraussetzung dafür, daß die Objektbezeichnungen als Referenz

verwendet werden können, ist, daß sie in Textform vorliegen und nicht nur als gesampelte Sprache.

- Um *räumliche Eigenschaften* - insbesondere die Ausrichtung von Gegenständen und ihre Anordnung in der Tiefe des Raumes - auszugeben, kann ebenfalls (synthetisierte) Sprache verwendet werden. Der Einsatz der Sprache zu diesem Zweck widerspricht allerdings der Forderung nach Medienkonstanz, also der durchgängigen Verwendung *eines* Mediums (Grafik), um *eine* Klasse von Informationen (geometrische Eigenschaften) darzustellen. Daher und weil natürlichsprachliche Orts- und Richtungsangaben oft nicht eindeutig sind, sollten gesprochene Hinweise über räumliche Eigenschaften bei der tastenden Erkundung von taktilen Grafiken vermieden werden.
- Als *Navigationshilfsmittel* bei der geleiteten Erkundung (Abschnitt 6.2) kann Sprache gut eingesetzt werden. Falls zur Navigationsunterstützung vor allem Richtungsangaben („links“, „rechts“, „Nord“, „Südwest“, ...) verwendet werden, können diese als Samples vorliegen und ausgegeben werden, da nur eine begrenzte Zahl von Ausgaben möglich ist, aus der je nach Interaktionskontext vom Navigationsunterstützungssystem die benötigte ausgewählt und wiedergegeben wird.

Spracheingabe wurde bisher für den Zugang Blinder zum Computer und seinen Anwendungen wenig untersucht, weil einerseits die vorhandenen Systeme noch zu unzuverlässig für den Alltagsgebrauch waren und andererseits keine Notwendigkeit für die Spracheingabe bestand, da Blinde Eingaben in der Regel über eine Tastatur vornehmen.

Beide Gründe für die Nichtbenutzung von Spracheingabe sind heute bei der Interaktion mit Grafiken nicht mehr stichhaltig: Einerseits funktioniert Spracheingabe heute zufriedenstellend [FELLBAUM 97], andererseits bedeutete jede Tastatureingabe für den Erfasser einer haptischen Grafik, daß er seine Hände von der Grafik weg und zur Tastatur hin bewegen mußte. Das kostet Zeit und bewirkt, daß bei der Rückkehr zur Zeichnung eine Neuorientierung erfolgen muß, wenn man an der Stelle weitermachen möchte, an der man zur Tastatur gewechselt hat.

Im einzelnen kann *Spracheingabe* zu folgenden Zwecken verwendet werden:

- *Kommandos* zur Steuerung der Erkundungsprogramme sind nur in geringer Anzahl vorhanden, weil der größte Teil der Interaktion direkt auf der Zeichnung mit dem Tablett erfolgt. Durch eine Anbindung der Spracheingabe an die Benutzungsoberfläche des Betriebssystems, die von den meisten Spracheingabesystemen unterstützt wird, können diese Kommandos auch per Sprache eingegeben werden. Da die Anzahl der Kommandos beschränkt ist, sind die Ergebnisse der Spracherkennung gut.
- *Anfragen* der Art „Wo ist <Objektname>?“ sind eine Kombination von Kommando („Wo ist“) und frei wählbarem Bezeichner. Das Kommando kann relativ leicht erkannt werden. Da der Objektbezeichner des gesuchten Objekts aus der Menge der in der jeweiligen Zeichnung vorhandenen Objekte stammen muß, ist auch hier der Suchraum begrenzt und die Erkennungsquote relativ hoch.
- Für die *Eingabe* von Objektnamen beim Zeichnen mit TDraw ist Spracheingabe hervorragend geeignet, weil gerade hier eine Unterbrechung des Zeichenvorgangs (zum Eingeben des Namens mit der Tastatur) diesen nachhaltig stören würde. Hier treten allerdings auch die größten Probleme bei der Erkennung auf, weil der zu erkennende Suchraum (Wortschatz) nicht eingeschränkt werden kann. Es ist nicht vorherzusagen, welche Objektbezeichnung der Benutzer für ein gezeichnetes grafisches Element wählt.

Zusammenfassend läßt sich feststellen, daß Sprache bei der Erkundung von Zeichnungen eine bedeutende Rolle spielt. Sie könnte dazu verwendet werden, die der Zeichnung zugrunde liegenden Informationen nicht-visuell zur Verfügung zu stellen. Wesentlich angemessener als diese konkurrierende Rolle ist allerdings die komplementäre Rolle der Sprache bei der Bereitstellung von grafisch nicht darstellbarer Information und als Kommunikationsmedium bei der Interaktion mit dem Anwendungssystem.

6.4 Synergieeffekte durch die Kombination mehrerer Medien

Sprachliche Kommunikation und Interaktion unterscheiden sich in vieler Hinsicht von bildlicher. Beiden gemeinsam ist ihr Zweck: die Informationsvermittlung. Die Sprache bietet jedoch ein höheres Abstraktionsniveau als das Bild. Dies gilt insbesondere für das gegenständlich abbildende Bild. Daraus ergeben sich Konsequenzen für die Beurteilung und den Einsatz von Sprache und/oder taktilem Grafik:

Wird in der sprachlichen Kommunikation eine Bezeichnung eines konkreten Gegenstands verwendet (z.B. „Tisch“), so bleibt es dem Empfänger der Nachricht überlassen, aus dem Begriff und seiner eigenen Erfahrungswelt eine konkrete Vorstellung zu schaffen. Das Bild eines Tisches hingegen nimmt dem Betrachter die Umsetzung des Begriffs in das konkrete Objekt ab. Das Beispiel „Tisch“ mag dies verdeutlichen. Der Begriff sagt nichts darüber aus, ob der Tisch einen, drei oder vier Beine hat und ob er rund oder eckig ist. Es bleibt dem Zuhörer/Leser überlassen, „sich ein Bild“ des genannten Tisches zu machen. Ein konkretes Bild eines Tisches nimmt dem Betrachter diese Arbeit ab. Damit wird er zwar entlastet, gleichzeitig aber auch in seiner Phantasie beschränkt. Es hängt vom Zweck der Kommunikation ab, ob der Tisch besser als „abstraktes Konzept“ oder als „konkretes Objekt“ kommuniziert werden soll.

Man stellt fest, daß sich sprachlich letztlich *nur* Konzepte vermitteln lassen. Diese können allerdings so detailliert sein, daß sie das Bild, das sich der Empfänger der Nachricht macht, wesentlich festlegen und ihm so wenig Spielraum für eine eigene Konkretisierung lassen.

Da sich diese Arbeit im wesentlichen mit *informierenden* Bildern beschäftigt, könnte man auf den Gedanken kommen, ganz auf Linien und andere tastbare Eigenschaften zu verzichten und die Szene nur durch ortsabhängige Sprachausgabe zu beschreiben: Dem blinden Erfasser könnte ein Tasttablett (ohne taktile Zeichnung) zur Verfügung gestellt werden, das auf Fingerdruck sprachlich ausgibt, was sich an der gewählten Stelle auf der Zeichnung befindet. Das Konzept (oder das „mentale Modell“) des dargestellten (aber nicht tastbaren) Gegenstands wäre so mit einem Ort verknüpft, und die Rekonstruktion der dargestellten Szene bliebe völlig dem Erfasser überlassen.

In der Tat wurden Versuche gemacht, die ein ähnliches Vorgehen zugrunde legen:

- Im Projekt „GUIB“ [GUIB 93] wurde ein 10 * 10 cm² großes Tasttablett in ein Dialoggerät integriert, mit dem Blinde den Bildschirm einer grafischen Benutzeroberfläche erkunden und so Fensterpositionen ermitteln konnten.
- In [Kennel 96b] wird ein System beschrieben, mit dem hierarchische Diagramme (z.B. Organigramme) ähnlich erkundet werden können.

Beide Ansätze haben zwei Aspekte gemeinsam:

Erstens handelt es sich bei den darzustellenden *Inhalten* um eher abstrakte Konzepte, die grafisch durch einfache Symbole (Rechtecke und gerade Linien) repräsentiert werden. Im einen Fall geht es um Fenster in einer Benutzeroberfläche, im anderen um das Verhältnis abstrakter Entitäten (z.B. Mitarbeiter einer Abteilung) zueinander. Die Visualisierungen

dieser Konzepte für Sehende kommen letztlich fast ohne zeichnerische Elemente aus: Allein die Position der Begriffe („Konzepte“) auf der Fläche spielt eine Rolle. Versuche, mit dieser Methode gegenständliche Darstellungen zugänglich zu machen, sind gescheitert³⁰.

Zweitens wurde in beiden Fällen zur Unterstützung der Orientierung auf der Fläche mit tastbaren Rastern experimentiert, die über das Tasttablett gelegt wurden. Diese taktile Unterstützung verbesserte die Erfassung der Positionen der Einzelelemente erheblich. Auch ohne Raster orientierten sich viele Benutzer an den Außenkanten des Tasttablets, sie verwendeten also ein Raster mit nur einer Zelle, um die Position der eigenen tastenden Finger relativ zu den verschiedenen Objekten und zum Rand der Darstellung besser bestimmen zu können.

Gerade diese letzte Beobachtung macht deutlich, daß Sprache und Bild sich ideal ergänzen, wenn es um das Erfassen von Konzepten geht, die in einer räumlichen Beziehung zueinander stehen. Bei gegenständlichen Bildern geht es nun nicht nur um die „Konzepte“ der dargestellten Gegenstände, sondern auch um deren konkrete Eigenschaften. Diese Eigenschaften lassen sich allerdings leichter erfassen, wenn man sie in einem konzeptuellen Zusammenhang wahrnimmt: Wird einem Erfasser beim Abtasten einer taktilen Darstellung eines Stuhls das Konzept „Stuhl“ sprachlich übermittelt, so hat er wesentlich weniger Probleme damit, die Einzelheiten des Stuhls (Anzahl der Beine, Form der Lehne, Ausrichtung im Raum usw.) zu erkennen, als wenn ihm diese konzeptuelle Information nicht gegeben wird.

Neben den in diesem Abschnitt bisher behandelten *multimedialen* Aspekten gilt es auch noch, auf *multimodale* Gesichtspunkte hinzuweisen: Es ist bekannt, daß Information besser erfaßt wird, wenn sie über mehr als einen Sinn wahrgenommen wird. In der Regel bedeutet die Einbeziehung anderer Sinne, daß auch andere Medien betroffen sind. Damit wäre jede *multimodale* Nachricht auch immer *multimedial*.

Text spielt hier als Medium eine Sonderrolle: Textuelle Nachrichten lassen sich sowohl verbal/akustisch als auch schriftlich/optisch übertragen, in Form von Braille-Schrift sogar taktil. Das Medium Text steht damit in mehreren Modalitäten zur Verfügung. In [DALY-JONES ET AL. 97] wird beschrieben, wie eine Kombination aus (schwarz auf weiß) geschriebenem Text und gesprochener Sprache einen besonders effektiven Nachrichtenaustausch ermöglicht. Die besondere Effektivität dieses multimodalen Ansatzes ist auch in der zusätzlichen Information begründet, die die Repräsentation des Textes in seiner jeweiligen Modalität mitliefert: Die *Tonlage* und Aussprache (vom Menschen) gesprochener Sprache ist oft im eigentlichen Wortsinn „vielsagend“. Das gleiche gilt für die *Handschrift* einer mit der Hand (und einem ggf. digitalen Stift) geschriebenen Nachricht.

Die Frage, die sich nun stellt, ist die nach der optimalen Kombination der verschiedenen Medien und Modalitäten. Wann soll gesprochene Sprache, wann sollen tastbare zeichnerische Elemente und wann gegebenenfalls nonverbale akustische Signale eingesetzt werden?

In der Realität ist der Raum für mögliche Antworten auf solche Fragen meist schon durch die zur Verfügung stehenden Ausgabegeräte und die vorhandene Datenbasis eingeschränkt: Steht keine Sprachausgabe zur Verfügung oder sind zu einer taktilen Zeichnung keine hinterlegten textuellen Beschreibungen vorhanden, so kann auch keine sprachliche Unterstützung stattfinden.

³⁰ Die diesbezügliche Begrenztheit der angesprochenen Vorgehensweise wurde von den Autoren von [GUIB 93] und [Kennel 96b] anlässlich von Konferenzen 1993 und 1996 bestätigt.

Beim Entwurf multimedialer, taktil grafik-gestützter Anwendungen für Blinde ist besonders auf die „Handhabbarkeit“ zu achten: Bei der Erkundung einer taktilen Grafik ist es sehr wichtig, daß die Hände immer auf der Grafik liegen können und nicht zu einem Braille-Display und von dort wieder zurück wechseln müssen. Ein solcher Wechsel unterbricht den Erfassungsvorgang, da sich die Hände erneut auf der Zeichnung orientieren müssen, bevor sie mit der Erkundung fortfahren können. In dieser Situation bietet sich per Sprachsynthese gesprochene Sprache als Modalität für das Medium Text an.

Auch für Eingabe von Kommandos und Daten während der Erkundung oder während des Zeichnens ist gesprochener Text (hier vom Rechner als *Eingabe* verwendet) das Medium der Wahl.

Da, wie beschrieben, taktile Zeichnungen weniger detailliert sein müssen als visuelle, lassen sich auf einer gegebenen Fläche, z.B. einem DIN A4 Blatt, möglicherweise nicht alle Informationen tastbar unterbringen. In diesem Fall kann man einzelne tastbare Elemente mit relativ viel textueller Information hinterlegen, die dann sprachlich ausgegeben wird.

Neben Sprache spielt auch die nonverbale Akustik in der Kombination mit tastbaren Elementen eine bedeutende Rolle. Akustische Signale können eingesetzt werden, wenn Hintergrundinformation zu einem komplexen oder großflächigen taktil dargestellten Gegenstand ausgegeben werden soll: Gehört eine Fläche z.B. zu einem Tisch, kann im Hintergrund das Geräusch einer über die Tischfläche streichenden Hand ausgegeben werden. Auf diese Art kann dem Problem begegnet werden, das viele Blinde beim Erkunden von Liniengrafiken haben: Die Bedeutung einer Linie ist nicht sofort erkennbar: Stellt sie einen länglichen Gegenstand dar, oder bildet sie die Grenze einer Fläche? Und wenn sie die Grenze einer Fläche darstellt, auf welcher Seite der Linie befindet sich die Fläche und auf welcher der „Hintergrund“? Durch akustische Signale, insbesondere durch nonverbale Hintergrundgeräusche, kann dem Erfasser zu jedem Zeitpunkt eine Art „Statusinformation“ mitgeteilt werden, die ihm diese Fragen beantwortet.

Technische Rahmenbedingungen schränken die Kombination taktiler und akustischer Informationsquellen schon in gewissem Maße ein. Das hier allerdings bedeutendere Kriterium ist die Zweckangemessenheit: Wie am Beginn dieses Abschnitts dargelegt, kommt es bei der Wahl des Ausgabemediums besonders darauf an, ob Konzepte oder räumliche Eigenschaften dargestellt werden sollen. Dem abstrakten Konzept ist das textuelle Medium Sprache angemessen, räumlich-geometrischen Eigenschaften konkreter Gegenstände wird die taktile Grafik eher gerecht.

Dieses Kapitel setzte sich mit verschiedenen Aspekten der geleiteten und selbständigen Erkundung von Grafiken auseinander. Während es für Sehende einige interessante und vielversprechende Ansätze zur Erkundungsunterstützung für visuelle Grafiken gibt (siehe z.B. [PREIM 98]), kann man auf dem Gebiet der tastbaren Grafiken für Blinde noch nicht auf einen großen Fundus an Methoden und Forschungsergebnissen zurückgreifen. Um nicht in den Bereich der Spekulation zu gelangen, wurde dieses Kapitel bewußt kurz gehalten und stützt sich nur auf wissenschaftlich abgesicherte Erkenntnisse.