

Grafisk representation av ljud

(HS-IKI-MD-04-305)

Mikael Lebram (a00mikle@student.his.se)
Institutionen för kommunikation och information
Högskolan i Skövde, Box 408
S-54128 Skövde, SWEDEN

Examensarbete på det datavetenskapliga programmet under
vårterminen 2004.

Handledare: Henrik Engström

Grafisk representation av ljud

Examensrapport inlämnad av Mikael Lebram till Högskolan i Skövde, för
Magisterexamen (M.Sc.) vid Institutionen för kommunikation och information.

2004-06-17

Härmed intygas att allt material i denna rapport, vilket inte är mitt eget, har blivit tydligt identifierat och att inget material är inkluderat som tidigare använts för erhållande av annan examen.

Signerat: _____

Grafisk representation av ljud

Mikael Lebram (a00mikle@student.his.se)

Sammanfattning

De senaste decennierna har en rad streckkodstekniker utvecklats för att grafiskt representera digital information. Gemensamt för dessa är en strävan efter hög densitet, det vill säga kapacitet att lagra mycket data på en liten yta. Densiteten begränsas emellertid av kravet på precision vid tolkning av den digitala datan, eftersom felaktigheter i digital data kan resultera i omfattande informationsförlust.

Analog information, exempelvis ljud, är däremot inte beroende av exakthet hos underliggande data i samma utsträckning. En analog ljudsignal kan beskrivas och återges approximativt, och ändå tolkas korrekt av en lyssnare. Detta faktum gör att det finns fördelar med att representera ett ljud utifrån dess analoga egenskaper. Exempelvis medför det sänkta kravet på absolut precision att densiteten hos representationen kan vara högre än hos konventionella streckkoder. I detta arbete utvecklas och utvärderas en grafisk representationsteknik som drar nytta av de analoga egenskaperna hos ljud – DAG (Dynamic Audio Graph). Ett flertal användningsområden för en sådan teknik har identifierats, bland annat musikdistribution via trycksaker, långtidsarkivering av ljud och illustration av ljud för upphovsrättsskydd.

Nyckelord: Streckkoder, grafiska representationer, datalagring, informationslagring, ljudmedia, ljudöverföring, DAG.

Innehåll

1	INTRODUKTION	1
2	BAKGRUND	3
2.1	ABSTRAKTIONSIVÅER HOS ANALOGA OCH DIGITALA REPRESENTATIONER	3
2.2	ANALOG REPRESENTATION AV LJUD	3
2.2.1	<i>Vinylskiva</i>	3
2.2.2	<i>Magnetband</i>	4
2.2.3	<i>Optisk film</i>	4
2.3	DIGITAL REPRESENTATION AV LJUD	5
2.4	DIGITAL REPRESENTATION AV BILDER	6
2.5	DIGITALA STRECKKODER	7
2.5.1	<i>Endimensionella streckkoder</i>	7
2.5.2	<i>Tvådimensionella streckkoder</i>	8
3	PROBLEMDEFINITION	11
3.1	DELMÅL	12
4	METOD	13
5	DAG – DYNAMIC AUDIO GRAPH	16
5.1	KODNING AV DAG	16
5.1.1	<i>Omsampling av ljuddata</i>	16
5.1.2	<i>Skapande av DAG</i>	18
5.1.3	<i>Utskrift av DAG</i>	21
5.2	AVKODNING AV DAG	22
5.2.1	<i>Inläsning</i>	22
5.2.2	<i>Identifiering och orientering</i>	23
5.2.3	<i>Tolkning och återskapande av ljud</i>	25
5.3	ANVÄNDARGRÄNSSNITT	27
6	PRAKTISKA DENSITETSTEST MED DIGITALA STRECKKODER	30
6.1	DATAMATRIX.....	30
6.1.1	<i>Hög skrivarupplösning och hög skannerupplösning</i>	31
6.1.2	<i>Hög skrivarupplösning och låg skannerupplösning</i>	31
6.1.3	<i>Låg skrivarupplösning och hög skannerupplösning</i>	32
6.2	PDF417.....	33
6.3	TESTANALYS	34
6.4	BERÄKNING AV DENSITET SOM CM ² PER LJUDSEKUND.....	35
7	DENSITETSTEST AV DAG	37
7.1	PILOTSTUDIE	37
7.2	LABORORIETEST MED FÖRSÖKSPERSONER.....	37
7.2.1	<i>Utformande av testmängd</i>	37
7.2.2	<i>Genomförande av test</i>	38
7.2.3	<i>Resultat</i>	39
7.2.4	<i>Kompletterande test av DAG med låg skrivarupplösning</i>	40
7.3	TESTANALYS	41
8	JÄMFÖRELSE MELLAN DAG OCH DIGITALA STRECKKODER	42
8.1	HÖG SKANNERUPPLÖSNING	42
8.2	LÅG SKANNERUPPLÖSNING.....	43
9	SLUTSATSER	44
9.1	PRÖVNING AV HYPOTES 1	44
9.2	PRÖVNING AV HYPOTES 2	44
9.3	ANVÄNDNINGSSOMRÅDEN FÖR DAG	45
9.3.1	<i>Verktyg för upphovsrättsskyddande av ljud</i>	45
9.3.2	<i>Artefakter</i>	45

9.3.3	<i>Distribuerade kvalitetsprov för skrivare</i>	45
9.4	RESULTATENS RELEVANS	46
9.5	FRAMTIDA ARBETEN	46
9.5.1	<i>Fortsatt utveckling av DAG</i>	46
9.5.2	<i>Kartläggning av kognitiva parametrar</i>	47
9.5.3	<i>Utforskning av digitala tekniker</i>	47
REFERENSER		48

1 Introduktion

Papper har under århundraden visat sig vara en tillförlitlig informationsbärare. De senaste årtiondena har pappret fått konkurrens av långt mer effektiva digitala medier såsom DVD, CD, hårddiskar, disketter och helt elektroniska minnesmoduler. Trots dessa högeffektiva innovationer är pappersmediet fortfarande använt i stor utsträckning vid förmedlande och lagring av information. Det är en billig form av materiell informationsbärare som även är enkel att distribuera i form av böcker, tidningar, etiketter och så vidare. Dessutom har papper en överlägsen tålighet mot mekaniska stötar, extern magnetism och temperatur jämfört med de flesta andra lagringsmedier, vilket gör mediet till en god kandidat för data- och informationslagring över lång tid (Smith, 1999).

Begreppen data och information används ofta synonymt trots att det finns väsentliga skillnader mellan dem. Data är faktauppgifter som i sig själva inte har någon mening, utan måste tolkas för att utgöra information (Avison et al., 1995). Information är resultatet från en tolkningsprocess av underliggande data, och är alltså per definition ingenting som kan lagras eller överföras. Trots detta används exempelvis termen *informationsöverföring* då vad som egentligen menas är *överföring av data som representerar information*. Detta arbete kommer bland annat belysa hur information kan representeras i olika former men ändå ha likvärdigt informationsinnehåll¹.

Digitala representationer av information har fördelen att kunna bearbetas och komprimeras med olika algoritmer, varigenom exempelvis mängden erfordrad data kan minimeras utan informationsförlust. Vid tolkning avkodas data till information som är identisk med den representerade. För att tolkning ska vara möjlig är det viktigt att all underliggande data kan tolkas korrekt, varför representationer av detta slag som regel erbjuder mekanismer för felkorrigering (Pavlidis, 2000).

Det finns en rad tekniker för att grafiskt representera digital information, det vill säga data som är avsedd för maskinell tolkning, på exempelvis papper (Pavlidis, 2000). De streckkoder som återfinns på så gott som alla produkter som säljs i handeln är exempel på sådana representationer. Streckkoder representerar som regel små datamängder som exempelvis produktbeteckningar vilka består av några siffror och bokstäver (Pavlidis et al., 1990). Det finns också mer komplexa tekniker som erbjuder större lagringskapacitet än vad som kan uppnås med konventionella streckkoder på samma yta. Med sådana tekniker kan godtycklig digital information representeras i grafisk form, exempelvis utförlig produktinformation, datafiler eller ljud (Wang, 1993).

Ljud är ett analogt fenomen som utgörs av tryckförändringar i luften (Rumsey, 1994), vilket innebär att en grafisk representation av ljud inte nödvändigtvis måste vara digital även om den är genererad av en dator. Syftet med detta arbete är att undersöka och påvisa de fördelar som finns med en representationsform som beskriver de analoga egenskaperna hos ljud, jämfört med existerande digitala representationsformer. En av de två hypoteserna som ställs är att en analog grafisk representation kan innehålla mer ljudinformation än en digital motsvarighet. Den

¹ I detta arbete kommer termen information att användas på det sätt gängse språkbruk bjuder, om skillnaderna inte är avgörande för vad som diskuteras.

1 Introduktion

andra hypotesen gäller att analoga representationer är mindre känsliga för hårdvaruvariationer än vad digitala är.

För att pröva de båda hypoteserna har en teknik för grafisk representation av analoga ljudsignaler utvecklats. Tekniken har därefter utvärderats med avseende på praktiska egenskaper såsom densitet och tolerans för hårdvaruvariationer. Dessa egenskaper har även undersökts i praktiken för två vanligt förekommande digitala representationstekniker. Resultaten från undersökningarna styrker den första hypotesen om att analoga grafiska representationer kan ha en högre densitet än vad som kan uppnås med digitala dito. Dessutom visar resultaten att den analoga tekniken är mer tolerant för hårdvaruvariationer vid inläsning än vad de digitala teknikerna är. En högre tolerans vid utskrift kan däremot inte påvisas, varför den andra hypotesen revideras. Hög densitet i kombination med flexibilitet vid inläsning gör emellertid att analoga ljudrepresentationer kan ha ett flertal tillämpningsområden:

- Den goda beständigheten hos papper och faktumet att analoga representationer kan tolkas utan digital elektronik är tilltalande för långtidsarkivering av språkliga dialekter och så kallade utrotningshotade ljud.
- Den höga lagringskapaciteten gör att ljud, exempelvis smakprov på musik, kan tryckas i tidskrifter på samma media som resten av innehållet. Flexibiliteten medger att representationerna kan tolkas utan specialhårdvara.
- Det unika mönster som karakteriserar ett visst ljud kan användas för att grafiskt illustrera ljud som ska upphovsrättsskyddas.

2 Bakgrund

Syftet med detta kapitel är en grundläggande redogörelse av de teorier och tekniker som leder fram till problemdefinitionen och utgör fundamenten för hypoteserna. Först ges en teoretisk orientering om analoga och digitala representationer. Därefter en översikt av tekniker för analog representation av ljud, följt av beskrivningar av tekniker rörande digitala representationer av ljud och bild. Sist i kapitlet redogörs för olika metoder för att grafiskt representera digital information.

2.1 *Abstraktionsnivåer hos analoga och digitala representationer*

En representation av ett fenomen kan sägas vara en beskrivning av relevanta aspekter hos det som ska representeras. Kravet som ställs på en representation är att det ska finnas någon form av funktionell relation mellan den representerade världen och den representerande världen (Palmer, 1978). Representationer som innehåller likvärdig information om den representerade världen kallas informationsmässigt ekvivalenta (Palmer, 1978), vilket innebär att formen på en representation inte spelar någon roll så länge den funktionella relationen bevaras. Exempelvis kan ett ljud representeras som elektrisk energi. Ljudet är då en del av representerade världen, och den elektriska energin utgör en representation i den representerande världen. Elektrisk energi kan representeras som en sekvens av reella tal. Reella tal kan representeras som en sekvens av bitar. Även om varje omvandling medför en viss informationsförlust bibehålls den funktionella relationen mellan det ljud som representeras och representationen på den högsta nivån, i det här fallet bitsekvensen. Digital information har fördelen att kunna omvandlas, komprimeras och krypteras förlustfritt. Detta innebär att representationer av digital information kan anpassas ur dessa aspekter med hjälp av olika algoritmer. Resultaten från varje omvandling är dock meningslösa utan kunskap om de underliggande algoritmernas eller abstraktionernas natur. Slutsatsen som kan dras av detta är att ju högre abstraktionsnivå en representation har, desto mer komplex blir tolkningsprocessen. Exempelvis har det skrivna ordet "häst" en högre abstraktionsnivå än en bild av en häst. Det skrivna ordet utgör en representation som kräver kunskaper om skrivet språk i allmänhet och det svenska språket i synnerhet för att vara tolkningsbart. Bilden däremot är en representation med egenskaper som står i mer direkt proportion till det representerade, varför tolkningsprocessen här förenklas (Stenning et al., 1995). Med avseende på komplexiteten hos tolkningsprocessen torde därför den universella användbarheten hos en representation öka ju mer analog den är med det representerade fenomenet.

2.2 *Analog representation av ljud*

Alla ljud örat kan uppfatta är förändringar av den omgivande luftens tryck mot trumhinnan. Dessa förändringar är analoga och kontinuerliga, vilket innebär att det finns oändligt många ljudfrekvenser inom det hörbara intervallet. Om ljudet fångas upp av en mikrofon genererar denna en elektrisk signal som står i proportion till ljudet. Det finns en mängd tekniker för att lagra denna analoga signal, för att vid ett senare tillfälle, mer eller mindre exakt, återskapa det ursprungliga ljudet. De mest utbredda teknikerna beskrivs nedan.

2.2.1 *Vinylskiva*

Vardera sidan på en vinylskiva innehåller ett sammanhängande spiralformat spår. Då ljud ska lagras på en skiva går den elektriska signalen som ljudet genererar till en liten

elektromagnet. Det magnetfält som skapas sätter ett gravyrstift i rörelse, samtidigt som skivan roterar. Gravyrstiftets rörelser resulterar i gropar och förhöjningar av olika storlek längs med spåret (Rumsey, 1994). När ljudet sedan ska spelas upp med en skivspelare är processen omvänd. Då skivan roterar löper skivspelarens nål genom spåret, varpå strukturen ger upphov till rörelser hos nålen. Dessa rörelser omvandlas via en elektromagnetisk spole till elektrisk ström som förstärks och går vidare till en högtalare, där elektriciteten omvandlas till hörbart ljud (Rumsey, 1994). De vanligaste² formaten på vinylskivor är LP (Long Play) och EP (Extended Play), med diametrarna 12 respektive 7 tum. Rotationshastigheten ska vara $33\frac{1}{3}$ respektive 45 varv per minut för korrekt återgivning.

2.2.2 Magnetband

Magnetbandstekniken använder ett plastband preparerat med magnetiskt påverkbara partiklar. Den elektriska signalen påverkar en elektromagnet i form av ett skrivhuvud, över vilket bandet löper. Magnetfältet som då skapas påverkar de magnetiska egenskaperna hos partiklarna på bandet. Vid avläsning löper bandet över ett läshuvud vilket fungerar på omvänt sätt, det vill säga att det genom induktion omvandlar magnetismen hos partiklarna till en elektrisk signal (Rumsey, 1994). Den vanligaste formen av magnetband är Compact Cassette, där bandet ska ha hastigheten $1\frac{7}{8}$ tum per sekund för att återge ljudet korrekt.

2.2.3 Optisk film

Optisk film är en av teknikerna som använts för att representera analoga ljudspår på rörlig film (Carr, 1998). Ljudspåret utgörs av en sekvens mörkare eller ljusare partier som löper parallellt med bildrutorna, och är alltså en del av den fotografiska filmremsan. När ljudspåret belyses släpper det igenom olika mycket ljus beroende på hur mörk eller ljus den belysta delen är. Detta ljus omvandlas till elektriska signaler som motsvarar ljusintensiteten med hjälp av en fotocell (se bild 1)(Carr, 1998; Rouse, 1930). De elektriska signalerna förstärks därefter och spelas upp som ljud i en högtalare. Ljudspårets hastighet är avhängigt den övriga filmens hastighet, vilken varierar mellan olika format.

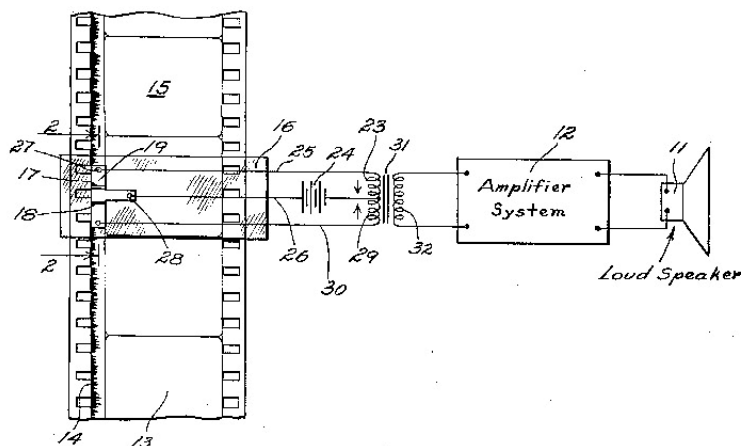


Bild 1: Skiss av fotocell för ljudrepresentation på fotografisk film (Rouse, 1930).

² En föregångare till vinylskivan var den så kallade stenkakan som roterade med en hastighet av 78 varv per minut. Vinylskivan ersatte så småningom stenkakan helt och hållet.

Denna teknik har med avseende på representationsformen fundamentala likheter med tekniken som kommer att beskrivas i detta arbete, men skiljer sig gällande fysiskt format och mediatyp.

2.3 Digital representation av ljud

För att ett ljud ska kunna lagras i en dator eller på en CD-skiva måste det först omvandlas till ett diskret digitalt format. En analog till digital signalomvandling (AD-omvandling) innebär att styrkan hos en kontinuerlig elektrisk signal approximeras till närmast liggande diskreta värde. Precisionen hos approximeringen bestäms av det antal binära bitar AD-omvandlaren använder för att beskriva de diskreta värdena. Antalet bitar brukar benämnas bitdjup, och en AD-omvandlare med bitdjupet 8 kan beskriva 2^8 olika värden. AD-omvandlaren mäter med jämna mellanrum amplituden (intensiteten) hos den elektriska signalen. Ju fler bitar AD-omvandlaren använder sig av, desto noggrannare kan amplituden bestämmas vilket medför ökad kvalitet (Rumsey, 1994).

En viktig faktor vid AD-omvandling av ljud är samplingsfrekvensen, som avgör hur ofta de ovan beskrivna amplitudmätningarna utförs. För att kunna beskriva en viss frekvens måste samplingsfrekvensen vara minst dubbelt så hög (enligt Nyquists teorem eller samplingsteomet (Roland-Mieszkowski et al., 2003)). Detta innebär att en korrekt återgivning av en ton med frekvensen 4000 Hz kräver en samplingsfrekvens av 8000 Hz, alltså 8000 olika amplitudvärden per sekund.

Alla amplitudvärden i ett insamplat ljud lagras som en sammanhängande sekvens av bitar. En generell okomprimerad ljudfil består dels av den bitsekvens som utgör själva ljuddata, dels information om hur ljuddata ska tolkas. Den information som krävs för att en bitsekvens ska kunna tolkas som ljud är:

- Bitdjup, det vill säga hur många bitar som används för att beskriva ett amplitudvärde. Typiska värden är 8 eller 16 bitar.
- Bitarnas inbördes struktur. För att tolka ett amplitudvärde som består av fler än en oktett måste ordningen på de representerade oktetterna vara korrekt. För 16-bitarsljud gäller som regel att oktetten med lägst signifikans representeras direkt framför oktetten med högst signifikans.
- Antal ljudkanaler, exempelvis 2 för stereo eller 1 för monoljud.
- Samplingsfrekvens, det vill säga hur många amplitudvärden som ska spelas upp per sekund. Typiska värden är 44100, 22050, 11025 eller 8000 Hz.

Ett 8-bitars monoljud med samplingsfrekvensen 11025 Hz läser 8 bitar av bitsekvensen 11025 gånger per sekund (se bild 2). Ett 16-bitars stereoljud med samplingsfrekvensen 44100 läser 16 bitar per kanal, alltså 32 bitar, 44100 gånger per sekund (se bild 3).

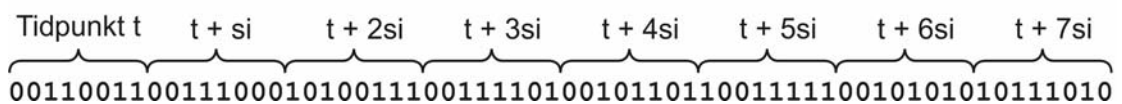


Bild 2: En bitsekvens tolkad som ett 8-bitars monoljud med samplingsfrekvensen s_f . Samplingsintervallet $s_i = 1/s_f$.

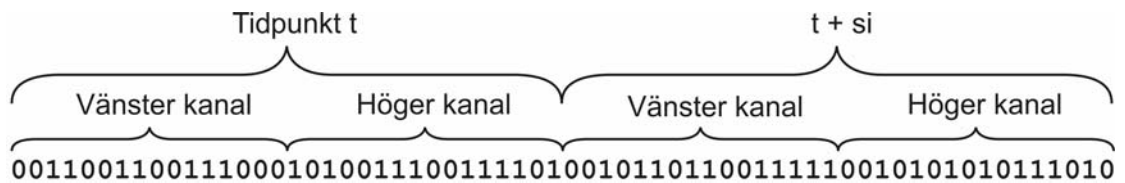


Bild 3: En bitsekvens tolkad som ett 16-bitars stereoljud med samplingsfrekvens f_s . Samplingsintervallet $s_i = 1/f_s$.

För att den digitala informationen ska kunna omvandlas till ljud igen måste den avkodas och genomgå digital till analog omvandling (DA-omvandling). Vid tolkning måste samtliga av de ovan nämnda parametrarna vara kända för att signalen från DA-omvandlaren ska likna originalsignalen. Signalen som DA-omvandlaren genererar är analog i bemärkelsen att den är kontinuerlig och motsvarar de rörelser den får en högtalare att utföra. Den är på grund av approximeringen däremot inte identisk med den analoga signal som digitaliserades, även om de båda signalerna kan vara mycket lika varandra.

2.4 Digital representation av bilder

Det vanligaste sättet att digitalisera en bild är att använda en flatbäddsskanner (Ohlson et al., 1999). Skannerns inläsningsmekanism består av en mängd ljuskänsliga sensorer som är jämnt fördelade över skannerns bredd. Varje sensor avger en elektrisk signal som står i proportion med den ljusintensitet den utsätts för (Burke, 1996). Genom att belysa bilden som ska skannas och successivt flytta raden av sensorer kan bilden omvandlas till ett digitalt format. Detta sker genom att signalen från varje sensor, i varje steg i längdriktningen, bearbetas av en AD-omvandlare. Ljusa partier reflekterar mer ljus än mörka, och genom att filtrera ljuset kan även färger tolkas (Ohlsson et al., 1999). Med ett bitdjup på 24 kan skannern göra skillnad på 256 toner av rött, grönt och blått och kombinera dessa till 256^3 (cirka 16,8 miljoner) olika nyanser. Skannerns optiska upplösning anger hur tätt sensorerna ligger, och avgör därför hur tätt två punkter kan ligga och ändå göras skillnad på. Detta värde mäts i enheten dpi (från engelskans "dots per inch", det vill säga punkter per tum) och ska inte förväxlas med pixlar, som är benämningen på de diskreta punkter som utgör den digitala bilden (Burke, 1996).

Det finns många olika typer av skrivare att välja bland för att få ut bilden i pappersformat igen, bland de vanligaste är dock bläckstråleskrivare och laserskrivare. Utskriftskvaliteten bestäms av skrivarens dpi-värde, vilket anger hur många punkter per tum skrivaren kan skriva. Detta gäller för både bläckstråleskrivare och laserskrivare. De båda typerna skiljer sig dock åt vid utskrift av gråskalor och färgnyanser (Cost, 1997). Gråskalorna som skapas av en laserskrivare består av ett rasterat mönster av svarta prickar mot vit bakgrund där förhållandet mellan svart och vitt motsvarar gråtonen i det aktuella området. Då en bläckstråleskrivare skriver ut gråskalor sker detta på liknande sätt, det vill säga att förhållandet mellan svarta och vita områden motsvarar den önskade gråtonen. Skillnaden är att bläckstråleskrivaren använder så kallade stokastiska raster, då den dispenserar färgdroppar i ett oregelbundet mönster. Tekniken vid färgutskriften motsvarar den för gråtoner för respektive skrivartyp och som regel används 3-6 olika färger som rasteras för att skapa olika färgnyanser (Cost, 1997).

2.5 Digitala streckkoder

Det finns ett antal tekniker för att representera digital information grafiskt, det vill säga som bilder vilka kan skrivas ut. Den utskrivna bilden kan sedan skannas in och tolkas för att återskapa den representerade informationen. Gemensamt för dessa tekniker är att de representerar diskret digital information vars exakthet och kompletthet är av stor vikt för användbarheten. Därför är kravet på tillförlitlighet stort, samtidigt som det ska vägas mot önskemål om hög densitet hos informationen, det vill säga att utnyttja mediet så effektivt som möjligt (Pavlidis, 2000).

2.5.1 Endimensionella streckkoder

Endimensionella streckkoder representerar information som en rad vertikala streck med olika reflektivitet, vanligtvis i svart och vit färg (Pavlidis et al., 1990). Denna typ av streckkod återfinns på nästan alla produkter som säljs i handeln, och används för att maskinellt läsa av exempelvis det nummer som identifierar en specifik produkt. Med detta nummer kan pris, namn, tillverkare och dylik information om produkten hämtas från en databas (Pavlidis et al., 1990). Koden kan avläsas genom att använda en laserskanner, fast eller handhållen, som låter en laserstråle svepa över strecken och registrerar hur mycket ljus som reflekteras.

För att vara så användbara som möjligt bör streckkoderna kunna avläsas från olika avstånd, i olika vinklar och med varierande hastighet (Pavlidis et al., 1990). Detta gör att strecken inte har en absolut storlek eller riktning ur skannerns perspektiv. För att hantera dessa problem översätts varje tecken av informationen som ska representeras till ett kodord, där varje kodord består av en bitsekvens som följer ett visst format. Det finns två grundläggande metoder att representera bitsekvenser som streckkoder, Width-code och Delta Code (Pavlidis et al., 1990).

För Width-code tilldelas varje bit ett streck eller ett mellanrum alternerande. Om biten är en etta blir strecket eller mellanrummet dubbelt så brett som det varit om biten varit en nolla. Ett exempel på Width-code är Code 39 vars respektive kodord består av nio delar (fem streck och fyra mellanrum), varav tre är breda (se bild 4).

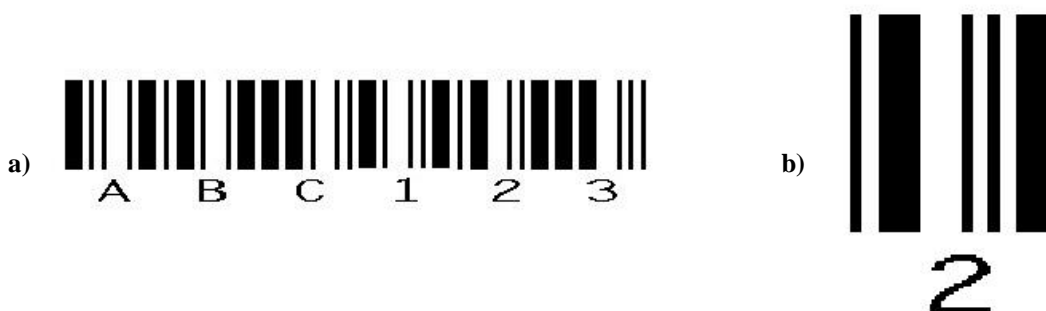


Bild 4: a) Width-code (Code 39). b) Detaljstudie av siffran 2 som representeras av bitsekvensen 001100001.

Med Delta Code delas kodorden in i moduler om en bit vardera. Moduler med ettor blir svarta och moduler med nollor blir vita, alltså mellanrum mellan strecken. Till de vanligare hör exempelvis UPC som har formatet (7,2), vilket innebär att ett kodord består av två streck och två mellanrum fördelade över sju moduler (se bild 5).



Bild 5: a) Delta Code (UPC). b) Detaljstudie av siffran 2 som representeras av bitsekvensen 0010011.

Utformningen av kodorden gör att representationerna blir självklockande, det vill säga att de absoluta värdena för storlek och avstånd inte har någon betydelse för tolkningen (Pavlidis et al., 1990). Dock medför detta begränsningar för hur många kombinationer som kan uppnås per kodord. UPC, vars kodord består av sju moduler vilka utan övriga begränsningar skulle kunna kombineras på 2^7 (=128) sätt, begränsas till 40 kombinationer på grund av att representationen ska bestå av två streck och två mellanrum. Code 39 kan på grund av dess restriktioner kombineras till 84 olika kodord.

Formatet på kodorden tillsammans med höjden på strecken gör att streckkoden kan läsas av i så gott som vilken vinkel som helst (Pavlidis et al., 1990). Däremot kan en felaktig avläsningsriktning orsaka problem, eftersom informationen blir helt annorlunda om den tolkas baklänges. I Code 39 har detta problem lösts genom att unika start- respektive stoppkoder ramar in representationen. UPC eliminerar problemet genom att inte tillåta att ett kodord kan spegelvändas och representera ett annat. Följden av denna restriktion är att det endast återstår 20 möjliga kombinationer av de tidigare 40. UPC används därför endast för representation av siffror, medan Code 39 även kan representera bokstäver (totalt används 44 symboler) (Pavlidis et al., 1990).

2.5.2 Tvådimensionella streckkoder

De endimensionella streckkoderna är effektiva för att lagra små datamängder, exempelvis ett produktnummer som refererar till en databas med information om produkten. Om databasen inte finns tillgänglig är streckkoden relativt meningslös eftersom den inte rymmer någon annan information. För lagring av data utöver ett produktnummer har ett antal tvådimensionella representationstekniker utvecklats, det vill säga tekniker som även utnyttjar höjdden av representationen. Den högre kapaciteten ger möjligheter att inte bara representera text, utan också godtycklig data såsom programfiler, ritningar och ljud. De tvådimensionella representationerna kan delas in i två huvudgrupper – stackade streckkoder och binära matriser (Pavlidis, 2000).

2.5.2.1 Stackade streckkoder

Stackade streckkoder innebär att flera endimensionella streckkoder staplas på varandra i höjddled. Ett exempel på denna teknik är PDF417, där PDF står för Portable Data File (se bild 6). Numret 417 anger att det är en Delta Code där varje kodord består av fyra streck och fyra mellanrum fördelade över 17 moduler (Wang, 1993).

2 Bakgrund

Tekniken fungerar i princip på samma sätt som UPC, med skillnaden att PDF417 inte bara måste skilja på kolumner utan även separera rader av kodord.



Bild 6: PDF417

Antalet kodord för PDF417 är begränsat till 929, där varje kodord kan representeras på tre olika sätt. Detta innebär att det finns tre varianter av kodordsuppsättningar, vilka benämns kluster, med vardera 929 olika kombinationer av streck och mellanrum. De tre klustren alterneras mellan raderna på så sätt att två vertikalt intilliggande kodord inte kommer från samma kluster (Wang, 1993). På detta sätt blir radseparationen självklockande.

För att tillåta viss snedhet vid inläsning kan höjden på varje element varieras mellan två till fem gånger bredden på elementen, varav det vanligaste förhållandet är tre gånger bredden (Pavlidis et al., 1990). Olika nivåer av felkorrigering kan väljas, vilket innebär att representationens egenskaper kan anpassas efter behov (Wang, 1993). En hög felkorrigeringsnivå ger hög tolerans för exempelvis skador på representationen (se bild 7), och kan därför vara lämplig för märkning av fraktgods och dylikt. En hög felkorrigeringsnivå medför dock att mängden redundant data ökar, vilket resulterar i att densiteten sänks. Den lägsta nivån av felkorrigering ger högre densitet men kräver att symbolen är intakt, varför denna nivå lämpar sig bäst i exempelvis kontorsmiljö.



Bild 7: Skadad PDF417

2.5.2.2 Binära matriser

Denna typ av tvådimensionella representationer är, jämfört med stackade dito, en mer direkt motsvarighet till den binära information som ska representeras. En binär matris består av ett rutnät där varje cell kan vara svart eller vit beroende på om det är en etta eller nolla som ska representeras. Ett exempel på binära matriser är DataMatrix (Priddy et al., 1990) (se bild 8).

2 Bakgrund

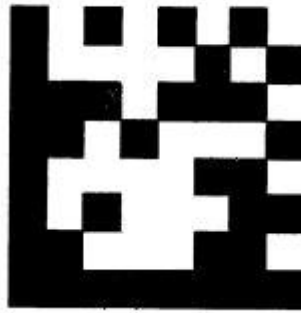


Bild 8: DataMatrix (Priddy et al., 1990)

En DataMatrix-symbols vänstra och nedre kant utgörs av heldragna linjer som används för att beräkna representationens area samt fungera som vinkelorientering vid inläsning (Priddy et al., 1990). Den högra respektive övre kanten utgörs av rutor som är alternerande svarta och vita, vilka används för att beräkna informationsinnehållet samt synkronisera tolkningen av matrisen (Priddy et al., 1990). Storleken och densiteten hos en representation kan således anpassas dynamiskt efter behov. Olika grader av felsäkerhet kan uppnås genom redundans, då samma data lagras på flera ställen över representationen.

3 Problemdefinition

Teknikerna för att grafiskt representera digital information har utvecklats med målet att de ska kunna representera mesta möjliga information på minsta möjliga yta. Essentiellt för dessa representationers användbarhet är att den diskreta digitala datan ska kunna tolkas helt felfritt, vilket är ett krav som medför begränsningar av densiteten (se kapitel 2.5). All typ av information är dock inte beroende av att underliggande data tolkas felfritt. Ljud och bild är exempel på sådan information, då det databortfall som orsakas av AD-omvandling inte behöver ha någon anmärkningsvärd påverkan av informationsinnehållet (se kapitel 2.3 och kapitel 2.4).

Genom att inte begränsas av krav på felfri underliggande data skulle ljudinformation kunna representeras grafiskt med en högre densitet än vad de i bakgrunden beskrivna teknikerna erbjuder. Representationer skapade med digitala tekniker ger dock inget stöd för approximativa tolkningar, det vill säga att den representerade informationen kan tolkas ungefärligt. Exempelvis kan ett felaktigt tolkat element i en DataMatrix-symbol resultera i dramatiska skillnader mellan representerad och tolkad data beroende på elementets signifikans i representationen (se kapitel 2.5.2.2). Om data representerades med färgnyanser istället för kombinationer av binära element skulle en feltolkning inte få samma proportioner, i synnerhet inte om den representerade informationen består av amplitudvärden för en ljudsignal³. Med bakgrund av detta resonemang ställs den första hypotesen detta arbete ska pröva: Ljudinformation kan representeras grafiskt med färgnyanser och på så sätt uppnå en högre densitet än vad som kan uppnås med binära grafiska representationstekniker. Denna hypotes kommer i arbetet att refereras till som hypotes 1.

En bild som representerar en amplitudskiftningarna hos en ljudsignal med färgnyanser skulle därför efter utskrift, på exempelvis papper, kunna skannas in och betraktas som en kontinuerlig analog signal som kan samplas in med godtycklig samplingsfrekvens och bitdjup. Insamplingsens noggrannhet skulle då, givet en ideal utskrift, endast begränsas av upplösningen och bitdjupet hos skannern. En låg skannerupplösning skulle resultera i en låg samplingsfrekvens och en hög skannerupplösning i en hög samplingsfrekvens. Ett lågt bitdjup skulle resultera i grövre approximationer än vad ett högt bitdjup skulle göra. De samplade ljudsignalerna skulle dock ha goda förutsättningar att vara informationsmässigt ekvivalenta med det representerade ljudet, om än med olika hög detaljåtergivning. Utifrån detta ställs en andra hypotes: En grafisk representation av en analog signal är mer flexibel med avseende på hårdvaruvariationer jämfört med digitala grafiska representationsformer. Denna hypotes kommer att refereras till som hypotes 2.

Utöver eventuell hög densitet och förmodad flexibilitet kan en analog grafisk representation av ljud motiveras utifrån dess låga abstraktionsnivå. Som beskrevs i kapitel 2.1 blir tolkningsprocessen för en representation mindre komplex ju mer direkta relationer representationen har med det som representeras. Ljusintensiteten

³ På grund av faktiska egenskaper hos skrivare och skannrar kan tolkningar av data som är representerad med nyanser *aldrig* väntas vara exakta, utan endast mer eller mindre noggranna approximationer. En representationsform som tolkas approximativt är därför inte lämpad att representera exempelvis text, eftersom en text är uppbyggd av diskreta tecken. Om ett "g" tolkas approximativt till något i alfabetet närliggande tecken, exempelvis "f" eller "i", kan detta få ödesdigra konsekvenser för tolkningen av texten som helhet.

3 Problemdefinition

hos nyanserna i en analog grafisk representation skulle stå i direkt proportion till den representerade ljudsignalens amplitud, vilket innebär att den enda information som krävs för tolkning är läshastighet och läsriktning. För tolkning av en digital grafisk representation av ljud krävs minst följande information:

- Hur den grafiska representationen ska avkodas till en bitsekvens.
- Hur många bitar i bitsekvensen som utgör ett amplitudvärde.
- I vilken ordning bitarna ska tolkas.
- Ljudets samplingsfrekvens.

Den mindre mängden information som krävs för att tolka en analog representation ökar sannolikheten att tolkning går att genomföra även utan kunskap om den underliggande tekniken. Applicerad på ett beständigt media, exempelvis papper, skulle en analog representation därför lämpa sig bättre för långtidsarkivering av ljud än en digital representation.

Syftet med detta arbete är att utveckla en teknik för grafisk representation av ljud, och utifrån den utvecklade tekniken pröva de båda hypoteserna.

3.1 Delmål

1. Utveckling av en teknik för att omvandla analoga ljudsignaler till analoga utskrifter och vice versa.
2. Analys av existerande digitala representationsteknikers praktiska prestanda. Det ena syftet med analysen är att erhålla jämförbara mått på hur hög densitet sådana representationer kan ha i praktiken, givet en viss hårdvara. Det andra syftet är att kartlägga hur olika upplösningar hos skrivare och skanner påverkar representationernas tolkningsbarhet.
3. Test av den utvecklade analoga representationsteknikens praktiska prestanda. Eftersom tolkning av ljud är en kognitiv process hos lyssnaren krävs ett antal lyssnare för att ge underlag för prövning av hypoteserna, det vill säga:
 - Att undersöka huruvida representationer skapade med den utvecklade tekniken kan ha högre densitet vad delmål 2 visar att digitala representationer kan ha.
 - Att undersöka huruvida representationer skapade med den utvecklade tekniken är mer toleranta för hårdvaruvariationer än vad delmål 2 visar att digitala representationer är.

4 Metod

Detta arbete kommer att bedrivas som en explorativ studie för att utveckla och utvärdera en teknik för grafisk representation av ljud. Studien kommer att spänna över ett flertal olika områden, alltifrån lågnivåegenskaper hos hårdvara till aspekter av mänsklig kognition via digital signalbehandling. Arbetet är uppdelat i delmål som vart och ett ska bidra till det huvudsakliga målet, det vill säga att pröva de båda hypoteserna utifrån den embryonala teknikens egenskaper. Det explorativa inslaget medför att metoderna som används för att uppnå delmålen i stor utsträckning kommer att bestå av praktiska experiment. Detta motiveras av arbetets tvärvetenskapliga karaktär och det faktum att både tid och utrymme är begränsat, vilket försvårar ingående studier av exempelvis fysisk struktur hos olika typer av media, kognitiva förmågor hos människan och olika algoritmer för signalbehandling.

Ett av arbetets syften är att jämföra lagringskapacitet hos representationstekniker som sinsemellan skiljer sig åt med avseende på representerad data. Streckkoder representerar godtycklig digital data, medan tekniken som ska utvecklas i detta arbete är ämnad att representera analog ljuddata. Denna skillnad medför olikheter i kvantifiering av representerad data. Kvantifiering av digital data görs som regel i termer av bytes eller bitar, medan analogt ljud enklast kvantifieras i termer av tid. För att en jämförelse ska vara möjlig krävs en homogenisering av kapacitetsmått. Detta kan uppnås genom att antingen beräkna hur många bytes ett digitaliserat analogt ljud upptar, eller genom att beräkna hur många sekunder ett digitalt ljud varar. På grund av arbetets fokus på analoga ljud kommer det senare alternativet att tillämpas.

På grund av att studien kräver testbara resultat kommer de ljud som behandlas att utgöras av talat språk. Det är relativt enkelt att avgöra huruvida ett talat budskap tolkas korrekt, jämfört med att bestämma om exempelvis ett musikstycke återges korrekt. Med informationsmässig ekvivalens menas därför i detta arbete att ett talat budskap tolkas lika av lyssnaren, oberoende av vilken representationsform som används⁴.

Delmål 1: Utveckling av en teknik för att omvandla analoga ljudsignaler till analoga utskrifter och vice versa

Den teknik som ska utvecklas kommer att bestå av fyra grundläggande delar:

- Simulering av en analog ljudsignal utifrån digital ljuddata.
- Omvandling av den simulerade signalen till nyansskiftningar i en utskrivbar grafisk representation.
- Inläsning av en analog signal bestående av nyansskiftningar i en skannad representation.
- Skapande av digital ljuddata genom sampling av den inlästa signalen.

Lämpliga algoritmer för bearbetning av ljud och bild kan delvis undersökas via litteraturstudier, delvis utvecklas genom experiment. Med lämpliga algoritmer menas sådana vars komplexitet står i proportion till resultatet. Exempelvis bedöms en funktion som tar 10 gånger så lång tid att exekvera som en annan funktion mindre lämplig om resultaten endast skiljer sig åt marginellt. På grund av en uttalad ovisshet om resultaten från varje moment är en detaljerad programdesign svår att åstadkomma

⁴ Därmed lämnas vissa rent subjektiva aspekter såsom ”ljudkvalitet” utanför detta arbetes ramar.

i förväg. Utvecklingen kommer därför att ske iterativt och utvärderas genom subjektiva bedömningar av i vilken mån tekniken är kapabel att representera ljudinformation.

För att kunna bearbeta digitalt ljud måste en undersökning av standardiserade ljudfilformat genomföras. Den typ av bearbetning som främst kommer att vara aktuell är interpolering av ljudfilens amplitudvärden, för att på så sätt simulera en kontinuerlig signal som kan samplas med godtycklig frekvens. Denna signal ska därefter representeras som nyansskiftningar i en bild. Nödvändig kunskap om filformat för ljud respektive bild kommer att hämtas från referensverk och online-specifikationer.

Förutom en informationsbärande del behöver representationen ha en layout som möjliggör och underlättar tolkning efter inskanning. Layouten kommer att utvecklas experimentellt, parallellt med algoritmerna för kodning och avkodning. Algoritmen för avkodning kommer även den att utvecklas experimentellt. Avkodningsalgoritmens syfte är att utifrån nyansskiftningar i den inskannade bilden simulera en kontinuerlig signal, som därefter samplas till en digital ljudfil.

Delmål 2: Analys av existerande digitala representationsteknikers praktiska prestanda

Det ena syftet med analysen är att erhålla jämförbara mått på hur hög densitet digitala streckkoder kan ha i praktiken, givet en viss hårdvara. Det andra syftet är att kartlägga hur olika upplösningar hos skrivare och skanner påverkar representationernas tolkningsbarhet. På grund av begränsade resurser är det inte möjligt att här utföra en uttömmande undersökning av samtliga förekommande streckkoder. Därför har två av de vanligast förekommande teknikerna valts ut: PDF417 och DataMatrix. Att dessa två hör till de vanligare är ett heuristiskt antagande grundat i antalet sökträffar på Internet⁵, och därför görs inga anspråk på att detta är helt överensstämmande med faktiska förhållanden.

Teoretiska egenskaper kommer att undersökas utifrån de tekniska specifikationerna för respektive teknik, det vill säga så kallade white-papers. Från sådana eller liknande skulle det också vara möjligt att skaffa uppgifter om vilka egenskaper de båda streckkoderna har i praktiken. En nackdel med denna metod är att det är en låg sannolikhet att respektive prestandauppgifter gäller identisk hårdvara, vilket kan försvåra en rättvis jämförelse. Därför kommer de båda teknikernas prestanda att undersökas genom praktiska experiment som utförs med en och samma hårdvara. Denna hårdvara ska också användas vid undersökning av prestanda hos den analoga representationsform som utvecklas i detta arbete. Experimenten väntas därigenom ge en grund för rättvisa jämförelser mellan de tre teknikerna.

Experimenten kommer att utföras genom att skriva ut och skanna in ett antal representationer skapade med de båda digitala teknikerna. De inskannade symbolerna ska därefter undersökas med avseende på tolkningsbarhet, som hos digitala representationer är relativt enkel att testa. Detta på grund av att som regel endast två

⁵ Sökningen gjordes med sökmotorn Google och söktermen bestod av namnet på aktuell teknik samt ordet "barcode". Förutom PDF417 och DataMatrix ingick följande 16 tekniker i undersökningen: Aztec code, ArrayTag, CodaBlock, Code 1, Code 16K, Code 49, CP code, Datastrip, Dot Code A, HueCode, MaxiCode, MiniCode, QR Code, SmartCode, Snowflake code och UltraCode.

utfall är möjliga, antingen går en symbol att tolka med intakt data eller så går den inte att tolka alls. Enkelheten möjliggör att en förhållandevis stor mängd streckkoder kan undersökas och varieras med avseende på fysisk storlek, skrivarupplösning och skannerupplösning.

Delmål 3: Test av den utvecklade analoga representationsteknikens praktiska prestanda

Med syfte att kartlägga tolkningsbarheten hos ljud som är genererade med den utvecklade representationstekniken kommer ett test med försökspersoner att genomföras. Testet ska utformas som ett interaktivt formulär, där en försöksperson själv spelar upp ett ljud med en knapptryckning och skriver in sin tolkning i en textruta.

Den höga graden av mänsklig inblandning gör att test av den analoga tekniken inte kan eller bör utföras på samma sätt som test av de digitala teknikerna. Detta motiveras av det faktum att en människa så småningom skulle lära sig hur testdatan skulle tolkas om den var likadan i samtliga tolkningsförsök. Det skulle vara teoretiskt möjligt att ha olika testdata i en stor mängd fall, men även då skulle resultatens relevans kunna ifrågasättas på grund av utmattning och dylika mänskliga aspekter hos de inblandade aktörerna (Kantner, 1994). Testfallen bör därför vara relativt få och skilja sig åt sinsemellan med avseende på innehåll. Med bakgrund av detta kommer testerna som utförs med den utvecklade tekniken i första hand syfta till att pröva hypoteserna som ställts i detta arbete. Resultaten väntas således inte ge ett exakt mått på representationsformens maximala lagringskapacitet.

De underliggande representationerna kommer att varieras med avseende på densitet, skannerupplösning och skrivarupplösning. Parametrarna kan komma att begränsas av antalet försökspersoner, då alltför många variationer i en liten testmängd kan försvåra analysen av resultaten (Kantner, 1994). Den parameter som har lägst prioritet bedöms vara skrivarupplösningen, eftersom skapandet av en representation är en engångshändelse medan inläsning kan ske vid flera tillfällen under olika förutsättningar.

För att erhålla riktlinjer om hur testmängden ska utformas kommer testet, som ska genomföras i laboratorium, att föregås av en pilotstudie. Testformulären i pilotstudien kommer att distribueras via e-post, vilket innebär att testdeltagarna själva får bestämma när och var de vill utföra testet. Fördelen med distribuerade testformulär är att studien kräver ett minimum av resurser, det vill säga att den är enkel att genomföra jämfört med ett laboratorietest. Nackdelen är att resultaten kan komma att påverkas av varierad miljö och hårdvara. För att reducera antalet parametrar kommer testfallen i pilotstudien därför endast att varieras med avseende på densitet.

Beräkningar av testresultaten kan göras på två sätt. Det ena sättet är att endast godkänna helt korrekt tolkade meningar och underkänna alla övriga tolkningar. Detta sätt att räkna kan liknas vid tolkningen av digitala representationer och lämpar sig därför bra vid en direkt jämförelse av tolkningsbarhet. Det andra sättet är att räkna andelen korrekt tolkade ord i meningarna, och på så sätt tillåta även delvis korrekt tolkad information. Detta sätt att räkna utnyttjar och påvisar det faktum att tolkning av analog information är mindre rigid än tolkning av digital information. Resultaten kan dessutom vara användbara vid en eventuell analys av huruvida enstaka ord påverkar tolkningen av hela meningar.

5 DAG – Dynamic Audio Graph

Tekniken för att representera ljudinformation som analoga bilder har givits arbetsnamnet DAG (Dynamic Audio Graph). I detta kapitel beskrivs dels de algoritmer som utvecklats för att skapa en bild från ett ljud och vice versa, dels arkitekturen som utgör själva representationen⁶.

5.1 Kodning av DAG

I denna del behandlas hur digital ljuddata tolkas och bearbetas för att kunna representeras av en DAG. Därefter beskrivs hur bilden som utgör en DAG byggs upp steg för steg, för att slutligen kunna skrivas ut på papper⁷.

5.1.1 Omsampling av ljuddata

Det första steget i ljud-till-bild-omvandlingen är att konvertera amplitudvärdena i ljudfilen som ska representeras till en sekvens⁸ av reella tal mellan -0,5 och 0,5.

Syftet är att underlätta omsampling, det vill säga öka eller minska samplingsfrekvensen. En ökning av samplingsfrekvensen påverkar inte informationsinnehållet, utan expanderar endast underliggande data. För att öka samplingsfrekvensen används här linjär interpolering, vilket innebär att ett amplitudvärde kan beräknas från en punkt som ligger mellan två ursprungliga samplingspunkter. Exempel: *Ett ljud med samplingsfrekvensen 10000 Hz består av en sekvens av amplitudvärden som spelas upp med 1/10000 sekunders intervall. En dubblering av samplingsfrekvensen till 20000 Hz medför att dubbelt så många amplitudvärden ska spelas upp med hälften så långt intervall (1/20000 sekunder). Detta innebär att ett amplitudvärde ska läggas till mellan varje par av amplitudvärden i den ursprungliga sekvensen (se bild 9). Sampling nummer 2 i det nya ljudet ligger mellan sampling nummer 1 och sampling nummer 2 i det ursprungliga ljudet (alltså sampling nr 1,5 i den ursprungliga sekvensen). Beräkningen av den mellanliggande amplituden sker enligt följande:*

us_i = amplitud för ursprunglig sampling i
 f = fraktion av heltal från önskat samplingsnummer
 Beräknad amplitud = $f(us_{i+1} - us_i) + us_i$
 I detta exempel: $0,5(0,2 - 0,0) + 0,0 = 0,1$

⁶ Tekniken bygger på en idé av Dr. Henrik Engström på högskolan i Skövde och är resultatet av en gemensam vidareutveckling och implementering av idén.

⁷ Andra media än papper är möjliga, exempelvis plast, plåt eller kopparlaminat, men en utvärdering av dessa ligger utanför ramarna för arbetet.

⁸ För närvarande representeras endast monoljud. Eventuellt stereoljud omvandlas till monoljud genom att beräkna medelvärdet på de båda kanalerna.

5 DAG – Dynamic Audio Graph

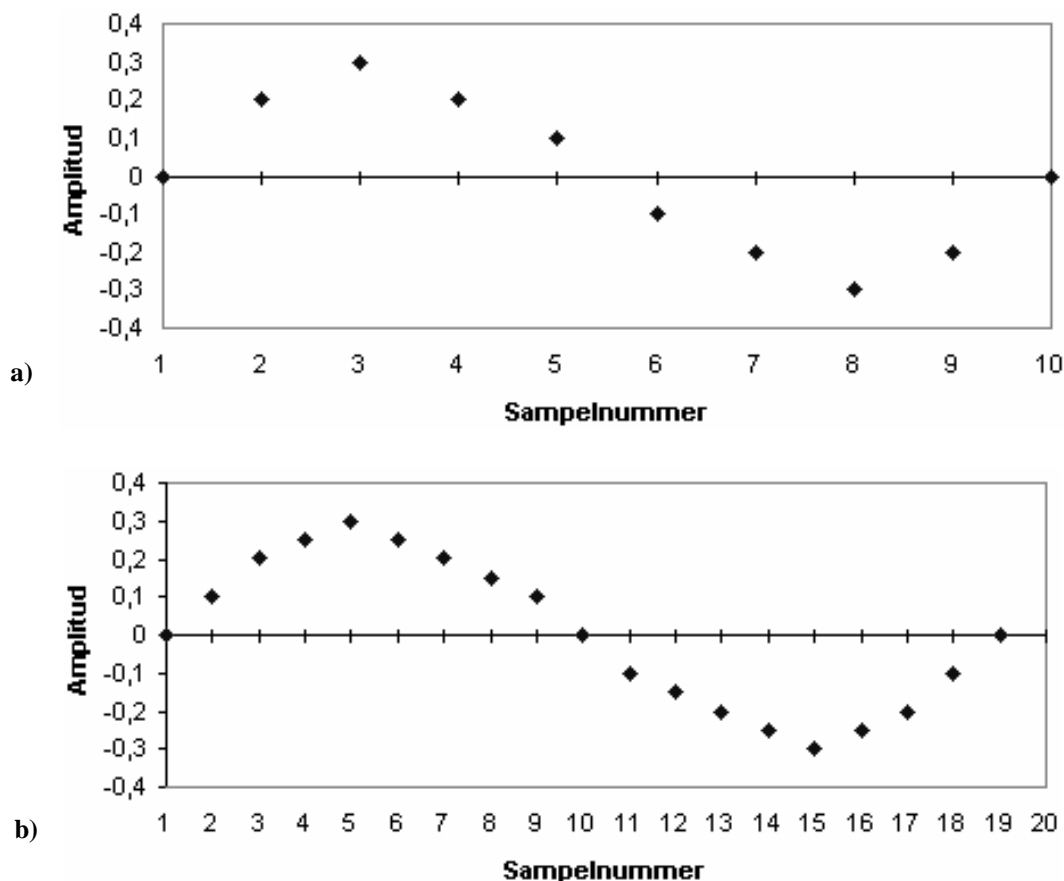


Bild 9: a) Amplitudvärden för den första millisekunden av ett ljud med samplingsfrekvensen 10000 Hz. b) En dubbling av samplingsfrekvensen hos ljudet resulterar i dubbelt så många amplitudvärden under samma tid.

En minskad samplingsfrekvens gör att underliggande data komprimeras och därmed kräver mindre lagringsutrymme än vad som krävs vid en högre samplingsfrekvens. Reduceringen av underliggande data medför dock en viss förlust av information. Linjär interpolering kan tillämpas vid minskning av samplingsfrekvensen på samma sätt som det tillämpas vid ökning. Det resulterande ljudet blir dock vasst och metalliskt på grund av ett fenomen som benämns aliasing (Story, 1997). Aliasing-problemet löses normalt genom att använda ett lågpas-filter som tar bort frekvenser högre än de som enligt Nykvists teorem (Story, 1997) går att representera med den lägre samplingsfrekvensen. För att undvika aliasing-problemet på ett enkelt sätt används i detta arbete inte interpolering vid minskning av samplingsfrekvensen, istället beräknas det nya amplitudvärdet som ett genomsnitt av de amplitudvärden som ska komprimeras. Metoden är primitiv, men har visat sig ge resultat som är tillfredställande nog för detta arbete. Exempel: *Ett ljud med samplingsfrekvensen 10000 Hz består av en sekvens av amplitudvärden som spelas upp med 1/10000 sekunders intervall. En halvering av samplingsfrekvensen till 5000 Hz medför att hälften så många amplitudvärden ska spelas upp med dubbelt så långt intervall, det vill säga med 1/5000 sekunders intervall. Detta innebär att ett amplitudvärde ska beräknas för varje par av amplitudvärden i den ursprungliga sekvensen (se bild 10). Sampling nummer 2 i det nya ljudet beräknas som genomsnittet av sampling nummer 3 och 4 i det ursprungliga ljudet enligt följande:*

5 DAG – Dynamic Audio Graph

us_i = amplitud för ursprunglig sampling nummer i
 n = antal ursprungliga amplitudvärden per nytt amplitudvärde
Beräknad amplitud = $(us_i + \dots + us_{i+n-1})/n$
I detta exempel: $(0,2 + 0,3)/2 = 0,25$

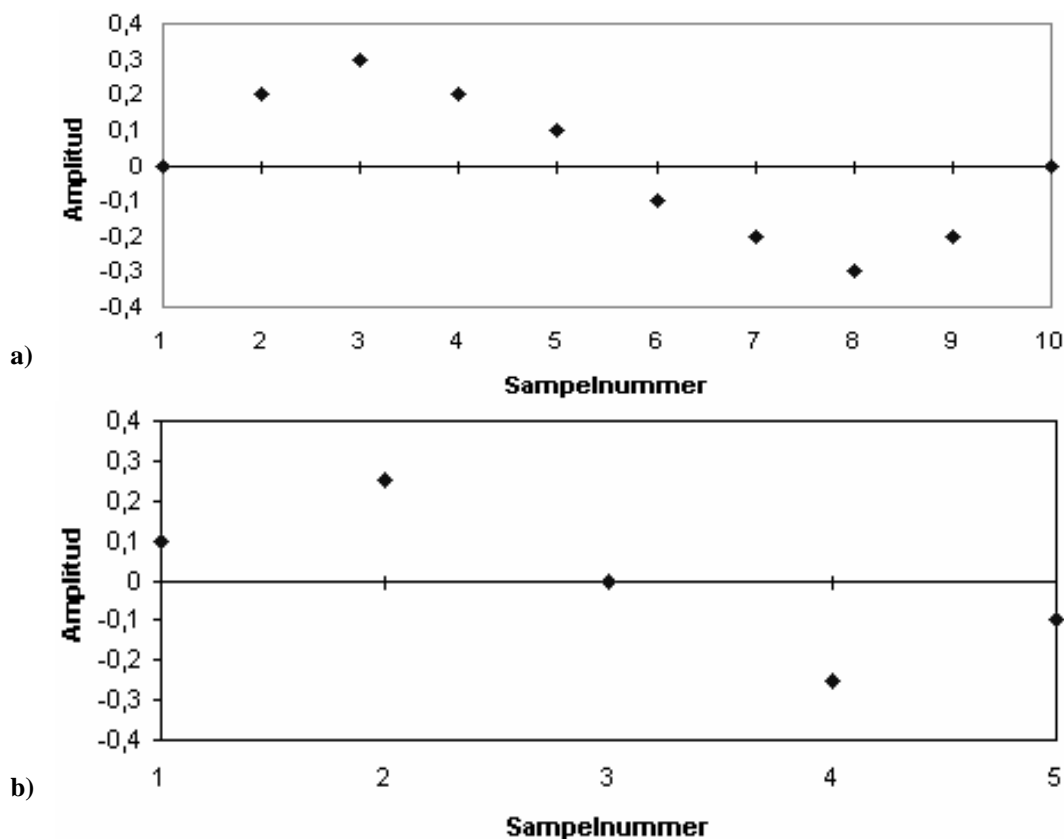


Bild 10: a) Amplitudvärden för den första millisekunden av ett ljud med samplingsfrekvensen 10000 Hz. b) En halvering av samplingsfrekvensen hos ljudet resulterar i hälften så många amplitudvärden under samma tid.

5.1.2 Skapande av DAG

En DAG är uppbyggd som en bitmap-bild, vilket innebär att den består av ett antal pixlar organiserade i en matris. Pixlarna kan tilldelas olika nyansvärden, beroende på vilket färgsystem som används (Sandoval et al., 1999). Det färgsystem som används i detta arbete är RGB, som representerar nyansen hos varje pixel med 24 bitar. Den första oktetten anger röda nyanser, den andra gröna och den tredje blå nyanser (se bild 11). Genom att kombinera olika värden av rött, grönt och blått kan 2^{24} nyanser representeras. Om samtliga tre oktetter tilldelas samma värde, erhålls en grå nyans, vilket ger RGB-färgsystemet kapacitet att representera 2^8 (=256) gråtoner⁹. För att en bitmap-bild ska visas korrekt krävs information om vilket färgsystem som används, bildens bredd och höjd i pixlar samt eventuellt storleken hos pixlarna¹⁰. Pixelstorleken bestämmer tillsammans med antalet pixlar på höjden och bredden hur stor yta bilden skrivs ut på.

⁹ Det förekommer även färgsystem som med hjälp av indexering kan representera 256 godtyckliga nyanser, även grå, med endast 8 bitar per pixel. RGB är dock en mer utbredd standard och ses därför som bäst lämpad för detta arbete, trots att 2/3 av kapaciteten inte utnyttjas.

¹⁰ Pixelstorleken är inte avgörande för att bilden ska visas med korrekta nyanser och proportioner på en bildskärm. Många enklare bildbehandlingsprogram tar inte hänsyn till parametern överhuvudtaget.

5 DAG – Dynamic Audio Graph

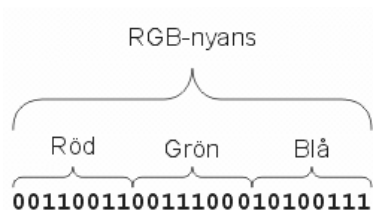


Bild 11: En 24-bitars RGB-nyans delas in i tre oktetter som anger värdena för rött, grönt respektive blått i nyansen.

För att representera ett ljud som gråtoner i en bitmap-bild måste samplingarna som görs omvandlas till RGB-nyanser. Konverteringen av ett amplitudvärde, som är ett reellt tal mellan -0,5 och 0,5, till en RGB-nyans som består av ett 24-bitars heltal, sker enligt följande:

$$\text{Oktett röd} = \text{Oktett grön} = \text{Oktett blå} = 255 * (\text{amplitud} + 0,5)$$

$$\text{Gråton} = \text{RGB}(\text{Oktett röd} \mid \text{Oktett grön} \mid \text{Oktett blå})$$

På detta sätt omvandlas amplitudvärden nära -0,5 till mörka gråtoner, och amplitudvärden nära 0,5 till ljusa gråtoner (se bild 12).

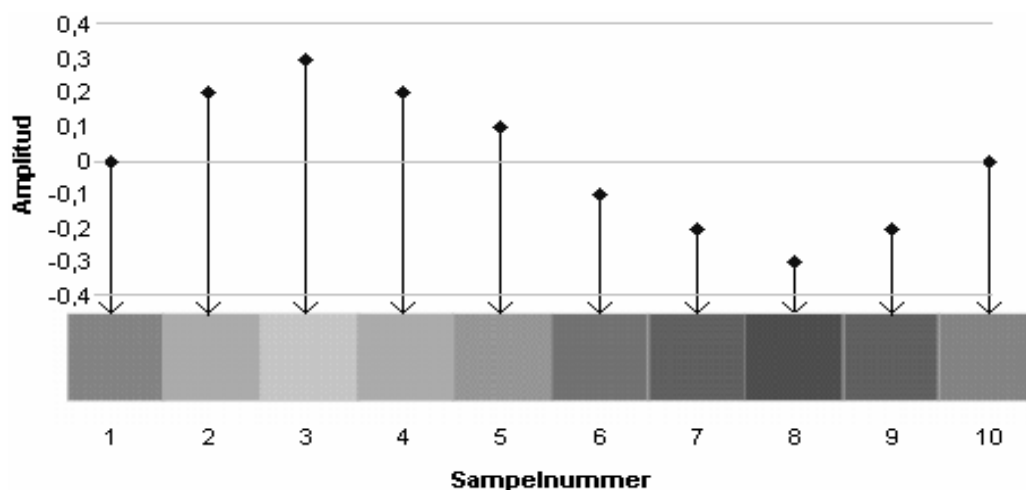


Bild 12: Amplitudvärden konverterade till gråtoner.

Storleken på den pixelmatris som utgör grunden för en DAG bestäms av följande parametrar:

- *Radlängd*, i cm. Med detta värde anger användaren hur bred representationen ska vara då den är utskriven på papper. Ju längre en rad är, desto fler pixlar får plats på den, och därmed krävs ett mindre antal rader för att rymma samtliga pixlar.
- *Radhöjd*, i pixlar. Varje sampling representeras med en och endast en pixel längs med raden. Genom att ange en radhöjd högre än 1 representeras varje sampling av flera pixlar staplade i höjdlid, och på så sätt skapas redundans i representationen. En representation med radhöjden 2 är dubbelt så hög som en representation med radhöjden 1.

Samt av två av följande tre parametrar:

- *Samplingsfrekvens, i Hz.* Här ges en möjlighet att öka eller minska samplingsfrekvensen med de metoder som beskrevs ovan. En ökad samplingsfrekvens ökar antalet amplitudvärden som ska representeras med pixlar, och därmed ökar också storleken på representationen. En minskad samplingsfrekvens minskar antalet amplitudvärden och resulterar därmed i en mindre storlek på representationen, men också i ett visst informationsbortfall.
- *Pixlar per cm.* Detta värde avgör hur många pixlar som får plats på en rad med angiven radlängd. Ju högre detta värde är, desto färre rader behövs av för att representera samtliga samplings, vilket i sin tur resulterar i en mindre storlek hos den utskrivna representationen.
- *Hastighet, i cm/sekund.* Detta värde anger hur många utskrivna radcentimeter som används för att representera en sekund av det representerade ljudet. Hastigheten kan också uttryckas som tid per rad.

Var och en av de sistnämnda tre parametrarna är beroende av de övriga två enligt följande:

$$\begin{aligned} \text{Samplingsfrekvens} &= \text{Pixlar per centimeter} * \text{Hastighet} \\ \text{Pixlar per centimeter} &= \text{Samplingsfrekvens} / \text{Hastighet} \\ \text{Hastighet} &= \text{Samplingsfrekvens} / \text{Pixlar per centimeter} \end{aligned}$$

Pixelmatrixens logiska mått, mätt i pixlar, beräknas utifrån ovanstående som:

$$\begin{aligned} \text{Totalt antal pixlar} &= \text{Antal amplitudvärden} * \text{Radhöjd} \\ \text{Antal pixlar per rad} &= \text{Bredd} * \text{Pixlar per centimeter} \\ \text{Antal rader} &= \text{Totalt antal pixlar} / \text{Antal pixlar per rad} \\ \text{Pixelmatrixens bredd} &= \text{Antal pixlar per rad} \\ \text{Pixelmatrixens höjd} &= \text{Antal rader} * \text{Radhöjd} \end{aligned}$$

Den utskrivna representationens fysiska mått, i centimeter, beräknas som:

$$\begin{aligned} \text{Bredd} &= \text{Pixelmatrixens bredd} / \text{Pixlar per centimeter} \\ \text{Höjd} &= \text{Pixelmatrixens höjd} / \text{Pixlar per centimeter} \end{aligned}$$

Vi kan nu skapa en pixelmatrix som innehåller tillräckligt många pixlar för att representera det aktuella ljudet med önskad samplingsfrekvens. Eftersom radhöjden är variabel är det inte helt korrekt att säga att varje sampling representeras av en *pixel*. Därför kommer begreppet *cell* att användas för att beskriva den eller de på höjden staplade pixlar som ska representera en sampling. Cellerna i DAG tilldelas nyanser radvis från vänster till höger, med början i den översta raden. Då slutet på en rad träffas på fortsätter tilldelningen av nyanser på nästa rad. När ljudet är slut, det vill säga när inga fler samplings kan göras, tilldelas återstoden av cellerna på den sista raden en gråton som representerar amplituden 0 (se bild 13).

5 DAG – Dynamic Audio Graph

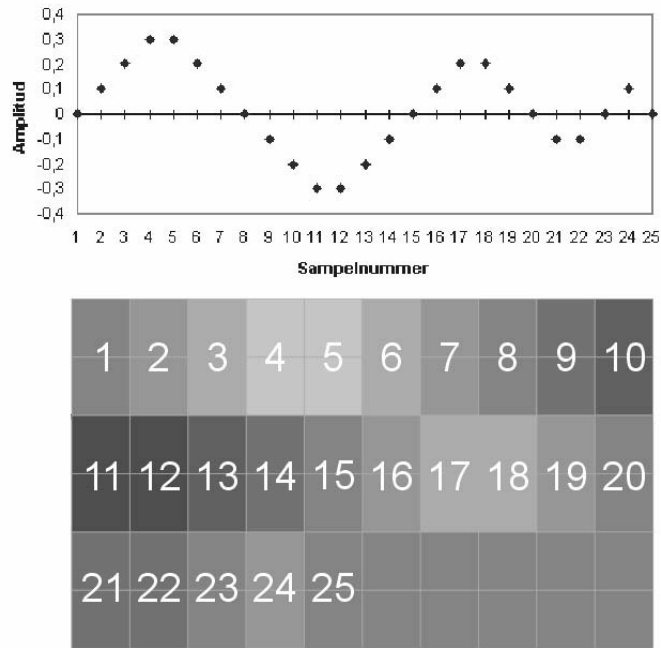


Bild 13: Ett ljud bestående av 25 samplingar konverterade från reella amplitudvärden (överst i bilden) till celler i en DAG (nederst i bilden). Radhöjden är 2, vilket innebär att varje cell är en pixel bred och två pixlar hög.

För att underlätta tolkning av DAG läggs två typer av orienteringshjälpmedel till i bitmap-bilden: Ett tak och ett antal ankare. Ett tak består av en svart rad, med samma längd och höjd som de övriga raderna, vilken placeras direkt över den första raden. Ett ankare har samma storlek som en cell och placeras till vänster respektive höger om rader med ojämnt ordningsnummer. Det horisontella avståndet mellan ett ankare och en rad är samma som den angivna radhöjden (se bild 14).

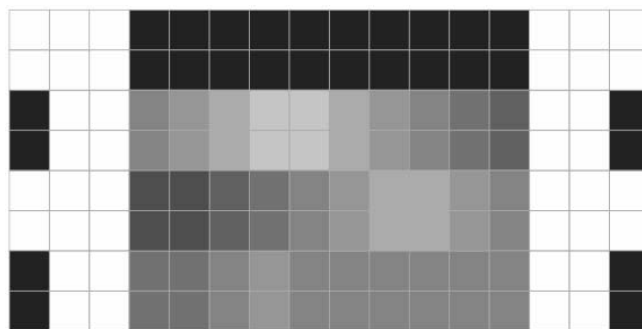


Bild 14: En komplett DAG med tak och ankare.

5.1.3 Utskrift av DAG

Som beskrevs i kapitel 2.4 skapar olika skrivartyper gråskalor på olika sätt, men att de som regel skapas med någon form av raster. Detta innebär att det i utskriften inte finns några egentliga grå nyanser, utan endast mönstrade fält med olika balans mellan vitt och svart (se bild 15).

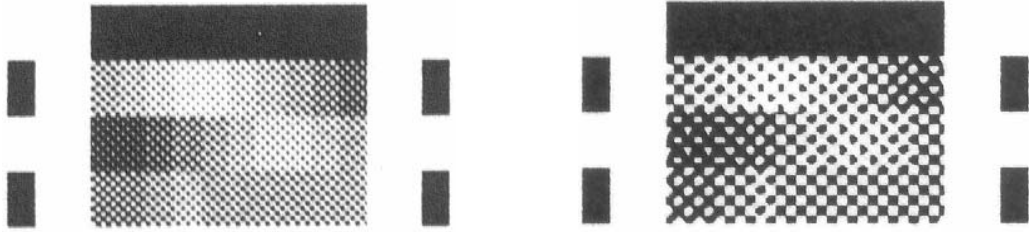


Bild 15: Grå nyanser skrivs ut som svart-vita mönster. Bilden till vänster är utskrivet med högre upplösning än bilden till höger.

5.2 Avkodning av DAG

Teoretiskt sett är det möjligt att med analog elektronik återskapa ett ljud som är representerat med en DAG. Ett sätt att göra detta är att låta en fotocell löpa genom representationens rader, och reproducera ljudet enligt principen för fotografisk film som beskrivs i kapitel 2.2.3. Dylka metoder medför dock praktiska utmaningar som ligger utanför detta arbetes ramar. Tolkningsprocessen som beskrivs här gäller därför DAG som är inläst och digitaliserad med hjälp av flatbäddsskanner¹¹.

5.2.1 Inläsning

Skannern läser in bilden som en bitmap-bild, där varje pixel tilldelas en grå nyans som motsvarar förhållandet mellan svart och vitt i det område som ska representeras av en pixel. Storleken på detta område avgörs av skannerns upplösning. Hög upplösning innebär många små pixlar (se bild 16), medan låg upplösning innebär få stora pixlar (se bild 17).

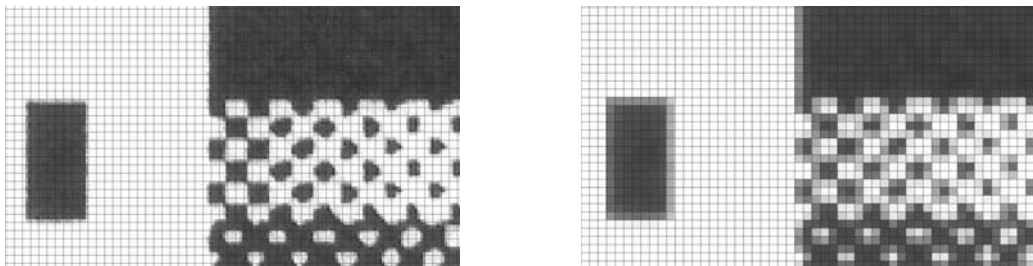


Bild 16: Detaljstudie av skanning med hög upplösning. De pixlar som delvis täcks av svart i bilden till vänster representeras som grå nyanser i skanningresultatet till höger.

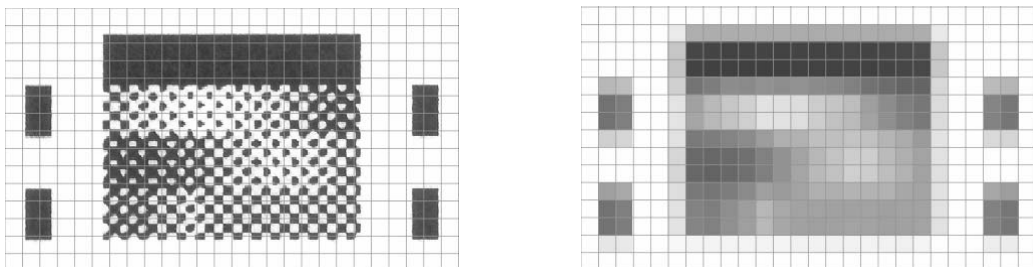


Bild 17: Skanning med låg upplösning. De pixlar som delvis täcks av svart i bilden till vänster representeras som grå nyanser i skanningresultatet till höger. De grå nyanserna motsvarar förhållandet mellan svart och vitt hos den inlästa bilden.

¹¹ Även andra digitaliseringsmetoder är möjliga, exempelvis via digitalkameror.

Ju högre skannerupplösningen är, desto mer noggrant återgiven blir den inlästa bilden. Värt att notera är dock att en låg upplösning i relativt stor utsträckning bibehåller representerade nyanser.

5.2.2 Identifiering och orientering

För att identifiera en DAG och dess rader i den inlästa bitmap-bilden eftersöks först representationens båda övre hörn. Sökningen sker genom en systematisk avläsning av nyanser hos pixlarna i bilden, uppifrån och ned - från mitten mot respektive ytterkant (se även bild 18):

```
x = bildbredd / 2
senasteTak = bildhöjd
LOOP WHILE NOT hittatHörn
  y = 012
  LOOP WHILE nyans(x, y) = ljus AND NOT hittatHörn
    y = y + 1
    hittatHörn = y > senasteTak + 5
  END LOOP
  IF NOT hittatHörn
    senasteTak = y
    x = x - 1
  END IF
END LOOP
hörnX = x
hörnY = senasteTak
```

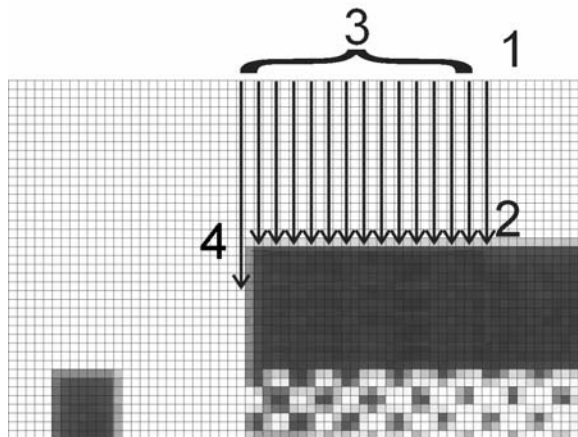


Bild 18: Sökning efter det övre vänstra hörnet hos en inläst DAG. 1 - Sökningen startar. 2 - En mörk nyans betyder att taket påträffats. 3 - Sökningen fortsätter stegvis åt vänster. 4 - En markant nivåskillnad betyder att taket är slut, och därmed har hörnet lokaliserats.

¹² Origo för pixelmatrisen i DAG är det övre vänstra hörnet.

5 DAG – Dynamic Audio Graph

Då hörnens positioner hittats eftersöks representationens ankare (se även bild 19):

```

x = 0
y = hörnY
stopX = hörnX
LOOP WHILE y < bildhöjd
  LOOP WHILE x < stopX AND nyans(x, y) = ljus
    x = x + 1
  END LOOP
  IF nyans(x, y) = mörk THEN
    minX = x
    maxX = x
    minY = y
    maxY = y
    i = 0
    LOOP WHILE i < 100
      LOOP WHILE nyans(x, y) = mörk
        IF x > maxX THEN maxX = x
        IF x < minX THEN minX = x
        IF y > maxY THEN maxY = y
        IF y < minY THEN minY = y
        x = x + SLUMP(-1 till 1)
        y = y + SLUMP(-1 till 1)
      END LOOP
      ankare = rektangel(minX, minY, maxX, maxY)
      i = i + 1
      nästaHörn = ankarhörn[i MODULU 4]
      x = nästaHörnX
      y = nästaHörnY
    END LOOP
    ankarvektor = ankarvektor + ankare
    y = maxY + 1
    stopX = maxX13
  END IF
  x = 0
  y = y + 1
END LOOP

```

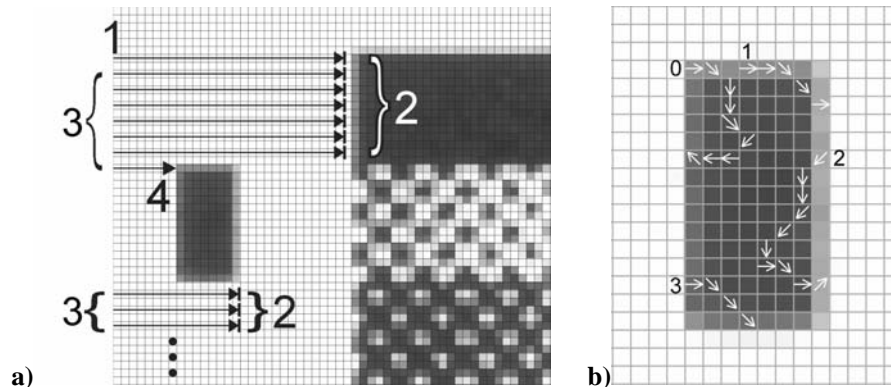


Bild 19: a) Identifiering av ankare. 1 - Sökningen startar. 2 - Sökningen stoppas eftersom inget ankare kan väntas längre till höger. 3 - Ny sökning börjar på nästa rad. 4 - En mörk nyans betyder att ett ankare påträffats. **b)** Ankarets utbredning utforskas slumpmässigt upprepade gånger. Varje ny rutt startar i något av det dithills kända ankarets hörn, med början i det övre vänstra hörnet (0), därefter alterneras hörnen medurs (1, 2 och 3). Observera att bilden endast beskriver fyra omstarter, i realiteten görs många fler.

När ankarna är kartlagda kan start- och slutposition för varje rad fastställas.

¹³ Sökningens maximala högerposition bestäms av senast funna ankarens högerkant. Därigenom blir sökningen mer tolerant för eventuell asymmetri i den inskannade bilden.

5.2.3 Tolkning och återskapande av ljud

Som visades i kapitel 5.2.1 kan en nyans i den ursprungliga bitmap-bilden breda ut sig över flera pixlar efter utskrift och inskanning. Därför är det inte lämpligt att tolka den inskannade bitmap-bildens nyanser pixel för pixel. Därför används ett *läshuvud*, som täcker flera pixlar, för att analysera nyanser i ett större område. Läshuvudet utgörs av ett antal segment som vart och ett är en pixel brett och ett angivet antal pixlar högt. Segmentens nyansvärde är genomsnittet av nyanserna i dess ingående pixlar. Läshuvudets nyansvärde kan erhållas på två sätt (se bild 20):

- Jämnt fördelad över hela området. I detta fall får läshuvudet det nyansvärde som motsvarar genomsnittet av samtliga ingående pixelnyanser.
- Avtagande påverkan från perifera segment. Här påverkar läshuvudets centrala segment dess nyansvärde mer än vad segment i ytterkanterna gör.

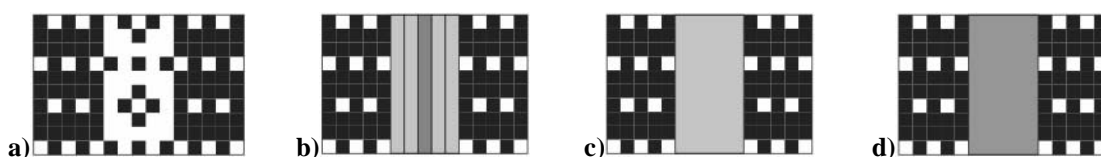


Bild 20: Nyanser i en pixelmatris erhålls genom att beräkna genomsnittet av nyanser i det aktuella området. **a)** En pixelmatris där pixlarna är antingen svarta eller vita. **b)** Fem läshuvudsegment med beräknade nyanser. **c)** Ett läshuvud vars nyansvärde har beräknats utifrån genomsnittet av samtliga nyanser i området. **d)** Ett läshuvud vars nyans påverkas mest av segmenten i mitten, och mindre av segmenten i ytterkanterna.

Egenskaperna för ett ljud som ska skapas från en DAG bestäms av följande parametrar:

- *Hastighet, i cm/sekund.* Detta värde anger hur många utskrivna radcentimeter som representerar en sekund av ljudet. För att ljudet ska återges med korrekt hastighet måste hastigheten som angavs vid skapandet av representationen vara densamma vid avläsningen. Även här kan hastigheten uttryckas som tid per rad.
- *Samplingsfrekvens, i Hz.* Samplingsfrekvensen anger hur långt avstånd det ska vara mellan läshuvudets avläsningar. Ett högt värde medför täta avläsningar, och ett lågt värde medför glesa avläsningar.
- *Längd på läshuvud, i pixlar.* Här anges hur många segment läshuvudet ska bestå av. Ett kort läshuvud är mer känsligt för små nyansvariationer än ett långt, vilket medför en ökad dynamik. Ett långt läshuvud resulterar å andra sidan inte i lika mycket distorsion i form av högfrekvent brus som ett kort läshuvud gör (se bild 21 och bild 22).
- *Höjd på läshuvud, i pixlar.* Detta värde anger hur många på höjden staplade pixlar ett läshuvudsegment består av. Ett lågt värde utnyttjar inte eventuell redundans i en rads höjdded. Ett värde som är större än höjden på en rad resulterar i överhörning, det vill säga att nyanser hos intilliggande rader tolkas som tillhörande den aktuella, vilket yttrar sig som ett eko i ljudet.

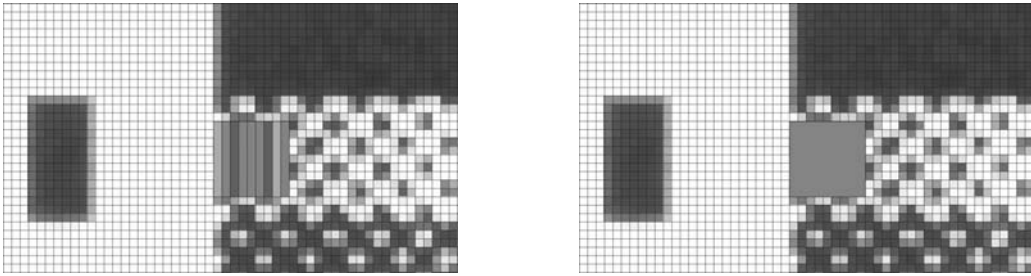


Bild 21: Ett läshuvud bestående av nio segment, vilkas genomsnittliga nyans ger läshuvudets värde.

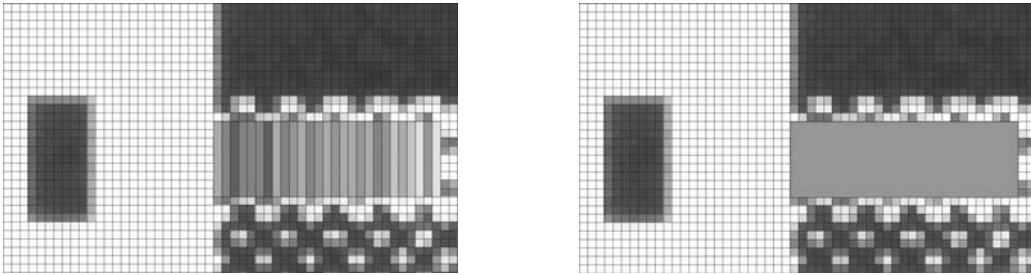


Bild 22: Ett läshuvud bestående av 27 segment, vilkas genomsnittliga nyans ger läshuvudets värde. Ju längre ett läshuvud är, desto mindre blir känsligheten för nyansvariationer.

Läshuvudet flyttas från vänster till höger, radvis uppifrån och ned. För att kompensera för eventuell snedhet hos den inskannade bilden används Bresenham's linjealgoritm för att flytta segmenten (Bresenham, 1965). När ett segment nått fram till slutet på en rad flyttas det ned till början av nästa. Antalet pixlar läshuvudet flyttas mellan varje avläsning bestäms av samplingsfrekvensen med en av följande metoder:

- Kontinuerlig steglängd, där steglängden är en funktion av hastigheten och angiven samplingsfrekvens, och är därför inte nödvändigtvis ett heltal. I den diskreta pixelmatrisen medför detta att det faktiska samplingsintervallet kan variera något.
- Diskret steglängd, där beräknad steglängd avrundas till närmsta heltal. Samplingsfrekvensen beräknas därefter om som en funktion av steglängden och hastigheten, och kan därför skilja sig från den angivna samplingsfrekvensen.

För varje steg läshuvudet förflyttas avläses dess nyansvärde, vilket konverteras till ett reellt amplitudvärde som läggs till i samplingssekvensen hos det ljud som skapas (se bild 23 och bild 24).

5 DAG – Dynamic Audio Graph

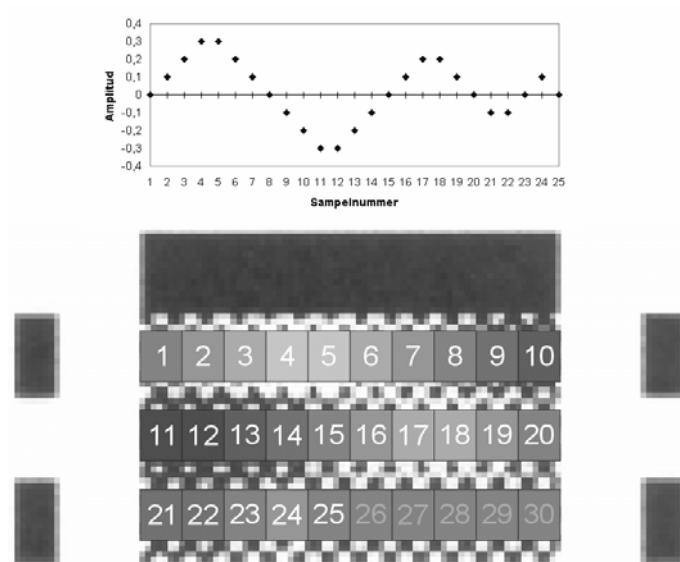


Bild 23: Läshuvudets nyansvärden konverteras till amplitudvärden i det representerade ljudet.

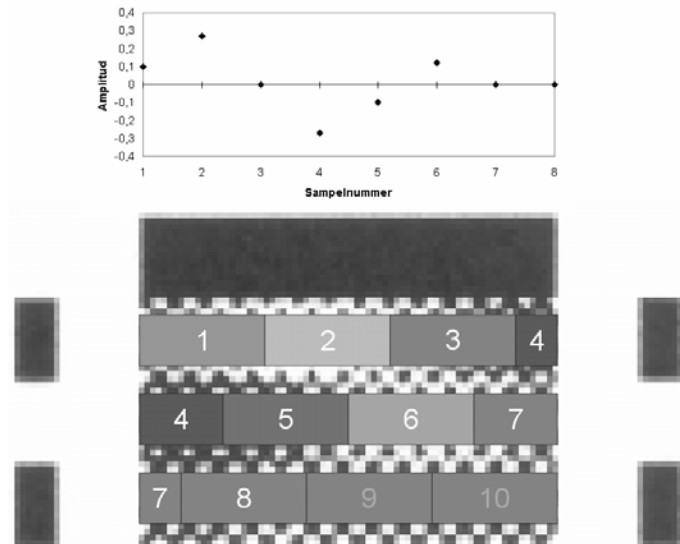


Bild 24: Läshuvudets nyansvärden konverteras till amplitudvärden i det representerade ljudet. Här är läshuvudet längre och samplingsfrekvensen lägre än i det ursprungliga ljudet. Trots detta bibehålls stora drag av ljudets egenskaper.

5.3 Användargränssnitt

För att förenkla skapande och tolkande av DAG utvecklades en applikation med ett fönsterorienterat användargränssnitt. En användare som önskar skapa en DAG med applikationen möts av delfönstret som visas i bild 25. Först och främst ska den ljudfil som ska representeras väljas. Därefter kan användaren välja vilken bredd den utskrivna representationen ska ha, samt den relativa radhöjden om extra redundans önskas.

Genom att markera en av variablerna *Frekvens*, *Upplösning* respektive *Hastighet* beräknas värdet för denna utifrån de båda andra. Om till exempel *Frekvens* är markerat kommer det valda ljudet att samlas med en frekvens som motsvarar produkten av de angivna värdena för *Upplösning* och *Hastighet*. För att illustrera

vilka effekter de olika värdena får på representationens utbredning förhandsgranskas den beräknade storleken i relation till ett A4-ark.

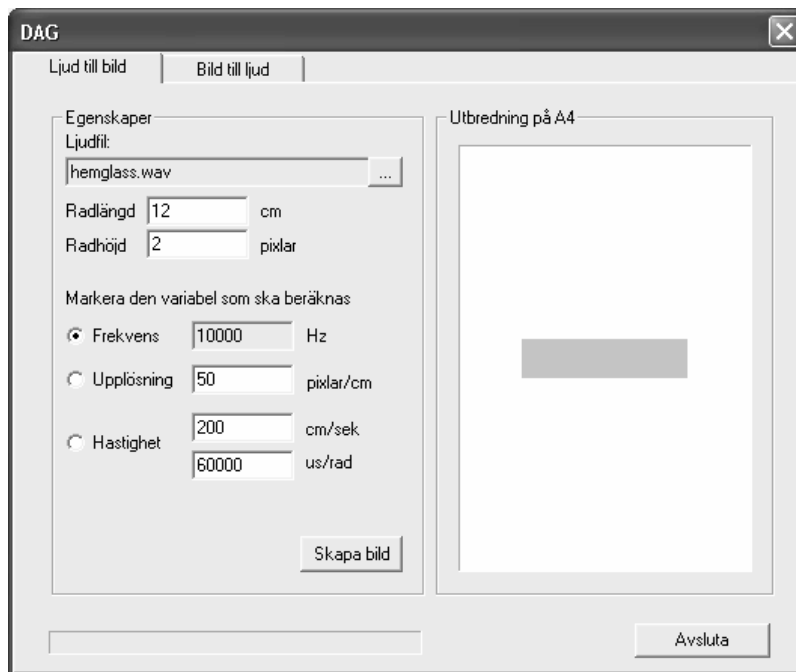


Bild 25: Användargränssnitt för skapande av DAG. Bilden visar exempel på inställningar för en DAG som representerar ljudet från en Hemglass-bil. Originalljudet samplas med frekvensen 10kHz. Om bilden skrivs ut med upplösningen 50 pixlar per cm blir bredden på representationen 12 cm.

Då användaren gjort de inställningar som önskas kan en DAG skapas i form av en bitmap-bild. De samplingspunkter som tas från det ursprungliga ljudet representeras som nyansskiftningar i bildens rader (se bild 26). Bilden sparas på disk för att senare kunna skrivas ut.

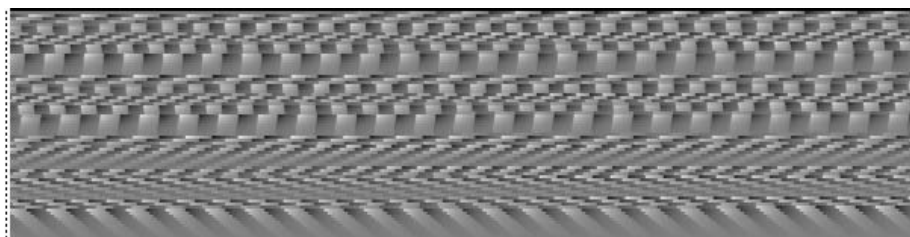


Bild 26: En DAG som representerar ljudet från en Hemglass-bil. Nyansskiftningarna i representationens rader motsvarar de samplingspunkter som tas från det ursprungliga ljudet. De perifera prickarna är ankarpunkter som ska vägleda kommande inläsning av DAG.

När en användare ska återskapa ett ljud från en inskannad DAG används delfönstret som visas i bild 27. Här kan användaren ange med vilken samplingsfrekvens DAG ska läsas in. Denna samplingsfrekvens är godtycklig, det vill säga det krävs ingen kännedom om vilken samplingsfrekvens som användes vid skapandet av bilden. Den enda parameter som måste vara känd är läshastigheten, det vill säga ljudtiden per rad.

Ljudets egenskaper kan justeras genom att variera storleken på läshuvudet, där ett kortare läshuvud resulterar i vassare ljud än vad ett långt gör. Ju längre läshuvudet är, desto mer dovt blir det resulterande ljudet. Även höjden på läshuvudet kan justeras.

5 DAG – Dynamic Audio Graph

Ett lågt värde utnyttjar inte redundant data till fullo, medan ett värde som är högre än radhöjden resulterar i överhörning. Dessutom kan metoden för att beräkna läshuvudets samplingsvärden väljas, liksom huruvida läshuvudets förflyttning ska beräknas kontinuerligt eller utföras med diskreta lika långa steg.

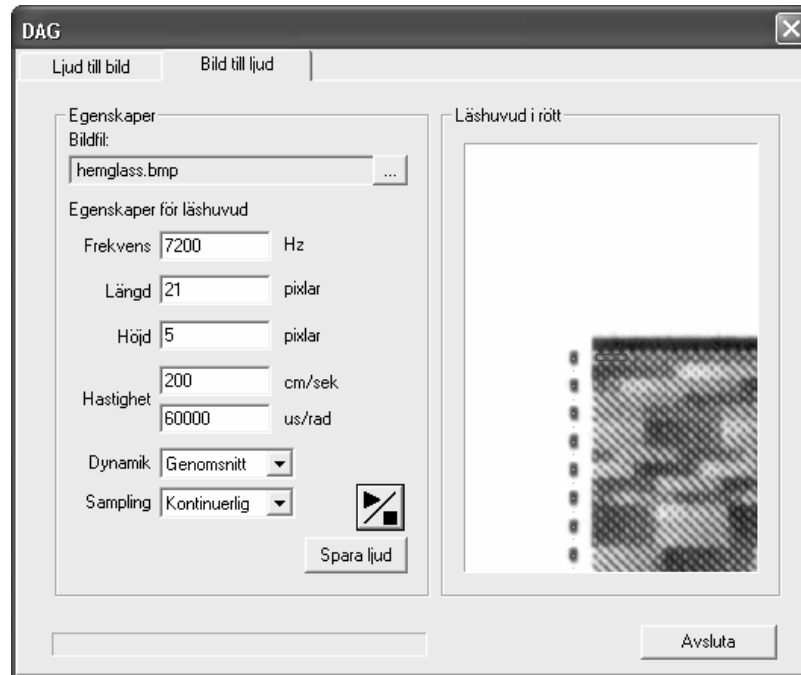


Bild 27: Användargränssnitt för återskapande av ljud. Bilden visar exempel på inställningar för återskapande av det ljud som representerades som en DAG i föregående bild. Observera att läshuvudets hastighet måste vara känd, medan samplingsfrekvensen kan anges med ett godtyckligt värde. Till höger i bilden visas läshuvudets storlek relativt representationens övre vänstra hörn, vilket utgör läshuvudets startposition.

När användaren har justerat parametrarna efter önskemål kan det återskapade ljudet sparas på disk som en digital ljudfil.

6 Praktiska densitetstest med digitala streckkoder

Den fysiska storleken hos en tvådimensionell representation avgörs av precisionen hos den utrustning som ska användas för att skriva respektive läsa den, och varierar därmed från fall till fall. I denna del redovisas experiment som genomförts med syfte att mäta maximal densitet, det vill säga antalet bytes som kan lagras per cm², hos DataMatrix (ISO 16022) respektive PDF417 (ISO 15438). Dessutom undersöktes vilka effekter förändringar av upplösning hos skanner och skrivare har på tolkningsbarheten för respektive symbol.

Den hårdvara som användes vid testerna var:

- Flatbäddsskanner Hewlett-Packard Scanjet 2200c, maximal optisk upplösning 600 dpi.
- Monokrom laserskrivare Brother HL-1250, maximal upplösning 1200 dpi.
- Paper av normal kontorskvalitet.

Mjukvaran som användes vid testerna var:

- RDataMatrix från J4L, för att skapa DataMatrix-symboler.
- PDFCenter från Symbol Technologies, för att skapa PDF417-symboler.
- ClearImage från Inlite, för tolkning av både DataMatrix och PDF417.

Den maximala lagringskapaciteten hos symbolerna begränsar mängden data som kan lagras per symbol. Data från flera symboler kan efter inläsning emellertid länkas samman så att de bildar en enda binärfil. PDF417 har inbyggt stöd för detta, till skillnad från DataMatrix, där sammanlänkningen måste göras externt. Eftersom möjligheterna till sammanlänkning finns, ses i detta arbete den maximala kapaciteten hos en symbol inte som en övre gräns för hur mycket data som kan lagras. Den beräknade densiteten antas därför gälla för en symbol med en oändlig fysisk yta¹⁴.

6.1 DataMatrix

En DataMatrix-symbol kan konfigureras på olika sätt, beroende på vilken typ av data som ska lagras samt hur stor mängd data som ska lagras (ISO 16022). De olika datatyperna innefattar tre olika varianter av text. Dessa skiljer sig åt sinsemellan med avseende på huruvida skillnad ska göras på versaler och gemener samt om specialtecken (@, \$, etc.) kan representeras. En fjärde datatyp är anpassad för representation av 8-bitars godtycklig data.

Den logiska storleken på en symbol avgör hur mycket data den kan lagra (ISO 16022). Den minsta symbolen består av 10x10 element och kan lagra enstaka tecken. Den största symbolen består av 144x144 element och kan lagra en text med upp till 3116 tecken, eller upp till 1556 bytes godtycklig data.

En serie experiment genomfördes för att avgöra vilken fysisk storlek som krävs för att en DataMatrix-symbol ska vara tolkningsbar med angiven hårdvara. Experimenten utfördes genom att en symbol skrevs ut fem gånger, varefter varje utskrift skannades in tre gånger. Eftersom arbetet fokuserar på ljuddata som sällan understiger 1556

¹⁴ Experiment har visat att ett visst avstånd krävs mellan två symboler för att de ska vara tolkningsbara var för sig. Storleken på detta avstånd minskar den totala densiteten hos de sammanlänkade symbolerna, men då detta avstånd inte är uppmätt i experimenten tas i detta arbete heller ingen hänsyn till den minskade densiteten.

bytes utfördes experimenten med en symbol konfigurerad till den fjärde datatypen och den största logiska storleken.

6.1.1 Hög skrivarupplösning och hög skannerupplösning

Syftet med experimentet var att undersöka hur hög densitet en DataMatrix-symbol kan ha med aktuell hårdvara.

1. En bild innehållande 11 symboler i olika storlekar skapades¹⁵. Symbolernas storlekar var fördelade med intervall om 1 mm, från 15 x 15 till 25 x 25 mm. Samtliga symboler var kompletta och logiskt identiska.
2. Bilden skrevs ut i fem exemplar med en skrivarupplösning på 1200 dpi.
3. Samtliga fem utskrifter skannades in tre gånger med en skannerupplösning på 600 dpi. Bilden flyttades något mellan varje inskanningstillfälle.
4. Tolkningsbarheten hos symbolerna i de inskannade bilderna kontrollerades med mjukvaran.

Totalt utfördes således 15 tolkningsförsök per symbolstorlek.

Resultat

Den minsta tolkningsbara symbolen som förekom i experimentet mätte 20 x 20 mm, vilket, eftersom symbolen representerar 1556 bytes, ger en densitet på cirka 390 bytes per cm². Symbolstorleken var tolkningsbar i 47 % av testfallen. Den minsta symbolstorleken som var stabil¹⁶ var 25 x 25 mm. Densiteten för denna symbol beräknas vara cirka 250 bytes per cm² med angiven hårdvara. Det bör tilläggas att denna symbolstorlek var tolkningsbar i samtliga testfall (se bild 28). Oregelbundenheten gällande tolkningsbarheten hos symboler med närliggande storlekar i såväl detta som efterföljande experiment behandlas i kapitel 6.3.

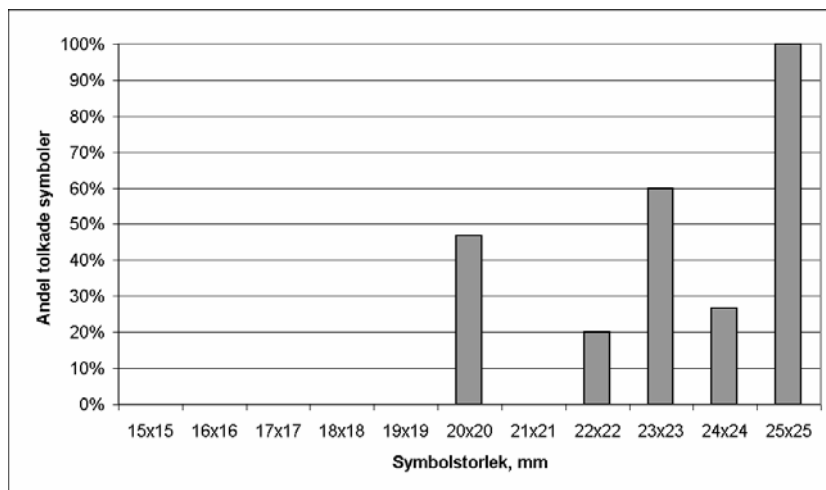


Bild 28: Tolkningsbarhet hos DataMatrix-symboler med olika storlekar, vilka är både utskrivna och inskannade med hög upplösning.

6.1.2 Hög skrivarupplösning och låg skannerupplösning

Syftet med detta experiment var att undersöka hur en låg skannerupplösning påverkar representationernas tolkningsbarhet. Ingen av symbolerna från föregående experiment var tolkningsbar, varför en ny serie symboler skapades:

¹⁵ Symbolstorlekarna bestämdes genom initiala experiment, där storleksintervallet var 5 mm. Den minsta förekommande tolkningsbara symbolstorleken var 20 x 20 mm. Därför valdes denna storlek som mellanstorlek i den nya testmängden.

¹⁶ En symbol anses här vara stabil om den är tolkningsbar i alla eller nära alla testfall.

6 Praktiska densitetstest med digitala streckkoder

1. En bild innehållande 11 symboler i olika storlekar skapades. Symbolernas storlekar var fördelade med intervall om 2 mm, från 30 x 30 till 50 x 50 mm. Samtliga symboler var kompletta och logiskt identiska.
2. Bilden skrevs ut i fem exemplar med en skrivarupplösning på 1200 dpi.
3. Samtliga fem utskrifter skannades in tre gånger med en skannerupplösning på 300 dpi. Bilden flyttades något mellan varje inskanningstillfälle.
4. Tolkningsbarheten hos symbolerna i de inskannade bilderna kontrollerades med mjukvaran.

Resultat

Halveringen av skannerupplösningen krävde att storleken på en stabilt tolkningsbar symbol var 40 x 40 mm (97 bytes per cm²), då 93 % av tolkningsförsöken lyckades. I 60 % av fallen kunde dock en symbol med storleken 32 x 32 mm (152 bytes per cm²) tolkas¹⁷ (se bild 29).

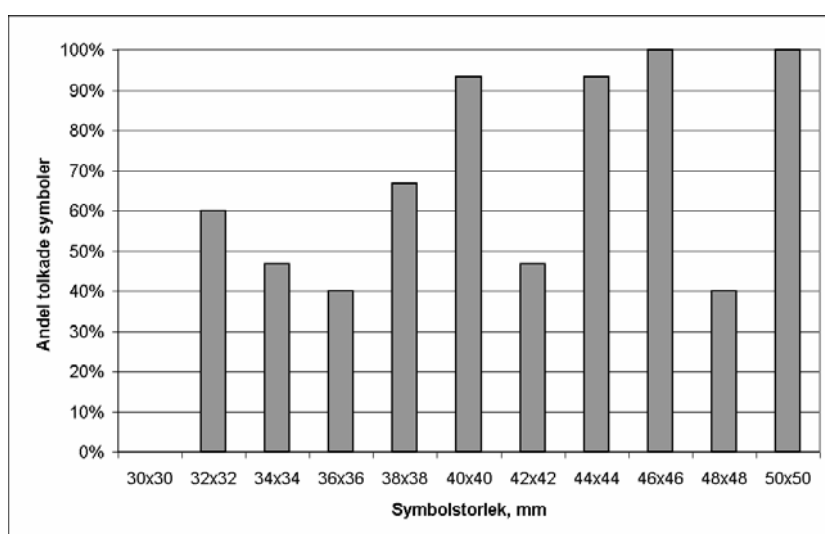


Bild 29: Tolkningsbarhet hos DataMatrix-symboler med olika storlekar, utskrivna med hög upplösning och inskannade med låg upplösning.

6.1.3 Låg skrivarupplösning och hög skannerupplösning

Syftet med detta experiment var att undersöka hur en låg skrivarupplösning påverkar representationernas tolkningsbarhet.

1. En bild innehållande 11 symboler i olika storlekar skapades. Symbolernas storlekar var fördelade med intervall om 2 mm, från 30 x 30 till 50 x 50 mm. Samtliga symboler var kompletta och logiskt identiska.
2. Bilden skrevs ut i fem exemplar med en skrivarupplösning på 300 dpi.
3. Samtliga fem utskrifter skannades in tre gånger med en skannerupplösning på 600 dpi. Bilden flyttades något mellan varje inskanningstillfälle.
4. Tolkningsbarheten hos symbolerna i de inskannade bilderna kontrollerades med mjukvaran.

¹⁷ Då testmängden kontrollerades med hög skannerupplösning (600 dpi) lyckades samtliga tolkningsförsök.

Resultat

Den lägre skrivarupplösningen medförde att symboler med små storlekar inte kunde tolkas. Den minsta symbolstorleken som kunde tolkas i förekommande fall var 32 x 32 mm (152 bytes per cm²). Symboler med storleken 36 x 36 mm (120 bytes per cm²) var tolkningsbara i samtliga 15 fall.

6.2 PDF417

En PDF417-symbol kan konfigureras olika beroende på vilken typ av data den ska innehålla samt vilken grad av felkorrigering som önskas. De olika datatyperna är numerisk data, text och 8-bitars godtycklig data (ISO 15438).

Den logiska storleken hos en PDF417-symbol avgörs av hur mycket data som ska lagras, graden av felkorrigering samt bredd-/höjdförhållandet hos elementen. Den maximala kapaciteten hos en PDF417 med lägsta nivån av felkorrigering är 2725 siffror, 1850 tecken i en text eller 1108 bytes godtycklig 8-bitars data (ISO 15438).

För att fastställa den minsta möjliga fysiska storleken hos en PDF417-symbol som är tolkningsbar med utrustningen som används i detta arbete genomfördes experiment motsvarande de som gjordes med DataMatrix. Experimenten gjordes med en symbol som lagrade 1108 bytes godtycklig data med den lägsta nivån av felkorrigering. Bredd-/höjdförhållandet hos elementen sattes till 1:2. Utifrån experimenten fastställdes den maximala densiteten hos en tillförlitligt tolkningsbar symbol till cirka 210 bytes per cm². Den minsta tolkningsbara symbolstorleken, 33 x 16 mm, var densamma för alla 15 fall i den första experimentserien (se bild 30).

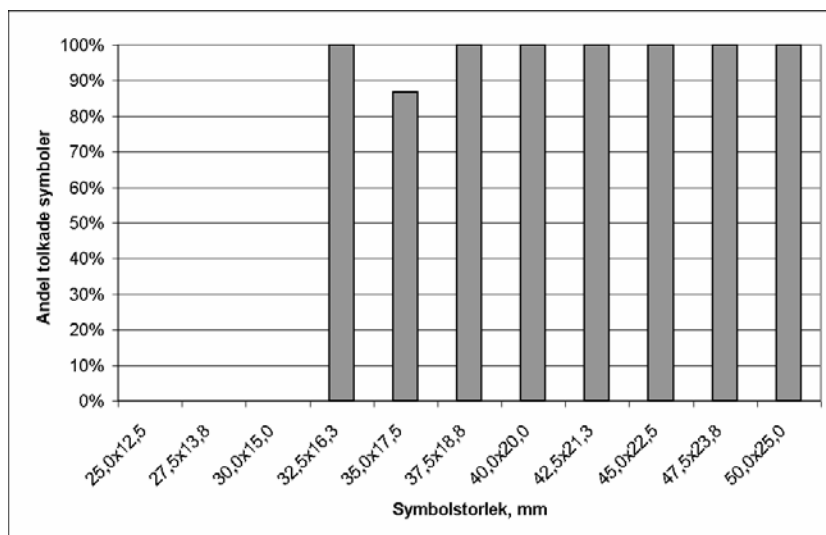


Bild 30: Tolkningsbarhet hos PDF417-symboler med olika storlekar, vilka är både utskrivna och inskannade med hög upplösning.

En halvering av skannerupplösningen krävde en symbolstorlek på 50 x 25 mm (89 bytes per cm²) för stabil tolkning. I 67 % av fallen kunde även symboler med måtten 47 x 23 mm (102 bytes per cm²) tolkas (se bild 31).

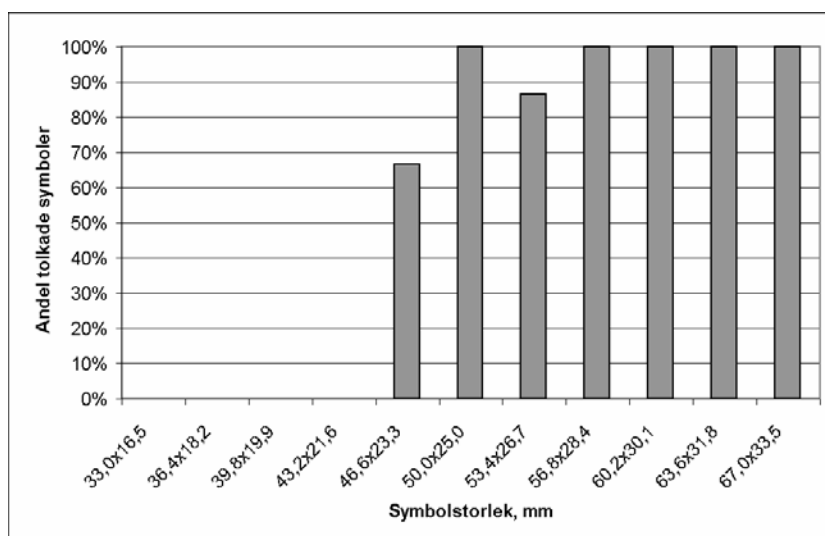


Bild 31: Tolkningsbarhet hos PDF417-symboler med olika storlekar, utskrivna med hög upplösning och inskannade med låg upplösning.

Vid experiment med låg skrivarupplösning och hög skannerupplösning var den minsta förekommande tolkningsbara symbolstorleken 43 x 21 mm (123 bytes per cm²). Denna storlek kunde tolkas i samtliga 15 fall.

6.3 Testanalys

Densiteten hos DataMatrix kunde i vissa fall uppnå 390 bytes per cm², det vill säga nära dubbelt så hög densitet som PDF417 (210 bytes per cm²). DataMatrix-symbolerna med högst densitet var dock instabila, stabilitet erhöles först vid en densitet på 250 bytes per cm². PDF417 uppvisade å andra sidan stabilitet genom samtliga experiment, skillnaden mellan minsta förekommande och minsta stabila symbolstorlek var som mest 6 %, jämfört med 43 % för DataMatrix.

En sänkning av skannerupplösningen medförde en lägre densitet hos båda symboltyperna. Detta kan förklaras med att skannern blir mer okänslig för detaljer, vilket resulterar i att de diskreta element som bygger upp en symbol inte kan isoleras lika lätt som vid en högre upplösning (se bild 32). Slutsatsen som kan dras av detta är att den höga densitet som kan uppnås med en hög skrivarupplösning kan göra en symbol otolkningsbar för en lågupplöst skanner.

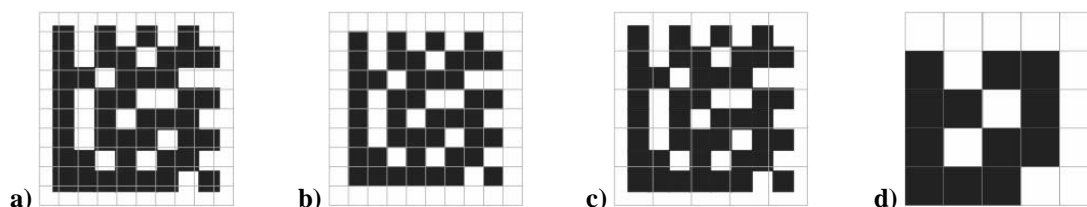


Bild 32: a) En utskrivna symbol ska läsas in med hög skannerupplösning. b) Resultatet av den högupplösta inläsningen motsvarar den utskrivna symbolen. c) Samma symbol ska läsas in med en lägre skannerupplösning. d) Resultatet av den lågupplösta inläsningen är inte tillräckligt detaljerat för att representera alla element i den ursprungliga symbolen.

En minskad skrivarupplösning innebar att densiteten minskade hos symbolerna, trots hög skannerupplösning. Orsaken är densamma som vid minskad skannerupplösning, med skillnaden att det nu är skrivaren som inte är tillräckligt detaljerad för att isolera

symbolernas element. Detta tyder på att skrivare och skanner måste ha liknande upplösningar om maximal verkningsgrad ska erhållas.

Det noterades även en viss oregelbundenhet vid tolkningsbarheten av närliggande symbolstorlekar, det vill säga att resultatet från tolkningen av en symbolstorlek var sämre än både större och mindre storlekar. Vad som orsakar fenomenet ligger utanför ramarna för detta arbete, men mycket tyder på att de olika storlekarna är mer eller mindre lämpade för specifika hårdvaruinställningar¹⁸. Utifrån de experiment som genomförts ges dock inga indikationer om att de uppmätta gränsvärdena för densitet skulle påverkas nämnvärt av att ytterligare justera representationernas fysiska storlek.

6.4 Beräkning av densitet som cm² per ljudsekund

Eftersom DAG enbart är avsedd för representation av ljud (som dessutom är approximativt tolkat) är enheten bytes per cm² mindre lämplig som jämförelsegrund av de olika representationstypernas densitetgränser. Ett mer lämpligt mått är att räkna antalet cm² som krävs för att representera en sekund av ett ljud. Denna del beskriver de antaganden som görs för att komma fram till ett sådant mått.

En beräkning av hur många bytes som krävs för att representera en sekund av ett ljud kan göras på ett antal olika sätt. Ett sätt är att räkna antalet bytes som används för att representera rådata, det vill säga ett obearbetat insamplat ljud. En sekund av ett ljud samplat med 16 bitar i 44 kHz upptar då 88 kB (16 bitar ger 2 byte x 44000)¹⁹. Om ljudet istället samplats med 8 bitar i 9 kHz upptar det 9 kB (8 bitar ger 1 byte x 9000). Eftersom digital data har fördelen att kunna komprimeras förlustfritt med hjälp av olika algoritmer gäller detta även för ljuddata. Denna typ av komprimering är dock ointressant för detta arbete eftersom det finns mer effektiva komprimeringsmetoder som är speciellt avsedda för ljuddata. Bland dessa hör MPEG-1 Layer III (ofta kallad mp3) (ISO 11172-3; ISO 13818-3) till de vanligast förekommande.

Komprimeringen som utförs av mp3 är inte förlustfri utan medför en reducering av mängden underliggande data (ISO 11172-3). Informationsförlusten står i proportion med komprimeringsgraden, det vill säga att ett ljud som komprimeras hårt förlorar mer information än ett ljud som inte komprimeras fullt så mycket. Tekniken är dock utvecklad med syfte att informationsförlusten ska upplevas som försumlig av lyssnaren, trots komprimering. Dataflödet, som vanligtvis benämns bitrate, anges i kbps (kilobits per sekund). Ju hårdaste komprimeringsgraden är desto lägre blir bitrate-värdet, det vill säga att det krävs mindre data för att representera ljudet. Den hårdaste komprimeringen standarden erbjuder resulterar i en bitrate av 8 kbps (ISO 13818-3). Typiska värden för musik är 128 eller 192 kbps, då kvalitén på ljudet anses vara jämförbar med ljud från CD-skivor. Även vid 8 kbps är ett ljud tolkningsbart, om än med kraftigt reducerad dynamik. Ett sekundlångt ljud som är komprimerat till 8 kbps upptar endast 1 kB (8000 bitar / 8). Den höga effektiviteten och det utbredda användandet av mp3 gör att denna standard kommer att användas som en måttstock i

¹⁸ I ett försök att finna orsaken till fenomenet användes formeln $F = \text{Rest}(R_x / (N_x / L_x))$, där R_x anger upplösning hos skrivare eller skanner och N_x anger antalet representationselement som är fördelade över en linje med längden L_x . Ett högt F -värde skulle med denna formel innebära att en större del av de representerade elementen går förlorade vid en diskretisering än om F -värdet är lågt. Ett visst samband mellan de påträffade tolkningssvårigheterna och höga F -värden har iakttagits men inte närmare analyserats, då resultaten från en ingående analys är av begränsad vikt för syftet med detta arbete.

¹⁹ Beräkningarna avser endast ljud med en kanal, det vill säga monoljud.

6 Praktiska densitetstest med digitala streckkoder

detta arbete. Det är dock inte sagt att det inte existerar än mer effektiva tekniker för komprimering av ljud, men en undersökning av hur detta förhåller sig ligger utanför arbetets ramar.

Testerna som utfördes tidigare i detta kapitel visade att DataMatrix som bäst kunde ha en densitet på 390 bytes per cm^2 . Det innebär att en symbol som representerar ett en sekund långt ljud, maximalt komprimerat med mp3, upptar en yta av cirka $2,6 \text{ cm}^2$ (1 kB / 390 bytes). Applicerat på den DataMatrix-symbol som var tolkningsbar i alla eller nära alla testfall skulle samma ljud uppta en yta av 4 cm^2 (1 kB / 250 bytes). Representerat med PDF417 skulle den erforderade ytan vara $4,8 \text{ cm}^2$ (1 kB / 210 bytes) (se tabell 1).

	Densitet	Erfordrad yta per sekund ljud	
		Rådata: 9 kHz, 8 bitar	mp3: 8 kbps
DataMatrix med maximal densitet (tolkningsbar i 47% av testfallen)	390 bytes per cm^2	23,1 cm^2	2,6 cm^2
DataMatrix, tolkningsbar i alla eller nära alla testfall	250 bytes per cm^2	36,0 cm^2	4,0 cm^2
PDF417 med maximal densitet (tolkningsbar i 100 % av testfallen)	210 bytes per cm^2	42,9 cm^2	4,8 cm^2

Tabell 1: Beräknad utbredning för olika ljudformat applicerade på digitala representationer.

7 Densitetstest av DAG

I detta kapitel undersöks praktiska egenskaper hos DAG för att skapa underlag för en jämförelse med streckkoderna som undersöktes i kapitel 6. På grund av att tolkning av ett ljud görs av en lyssnare och inte av algoritmer genomförs detta test med hjälp av försökspersoner. Ljuden som används för testning utgörs av meningar som är upplästa på svenska.

7.1 Pilotstudie

För att erhålla riktlinjer om hur ett test med försökspersoner skulle utformas genomfördes en pilotstudie. Ljuden i pilotstudien var indelade i tre olika klasser som skilde sig åt med avseende på densitet:

- Klass 1: 6,2 cm² per ljudsekund
- Klass 2: 3,2 cm² per ljudsekund
- Klass 3: 1,2 cm² per ljudsekund

Ljuden bestod av meningar upplästa av tre olika röster, två män och en kvinna. Varje röst förekom minst en gång i var och en av klasserna. Totalt bestod testmängden av 10 ljud. Ett interaktivt testformulär skapades, där försökspersonen lyssnade på ett ljud genom att klicka på en knapp. Därefter skulle de uppfattade orden skrivas in i en textruta. Testformuläret skickades via e-post till 15 försökspersoner, vilka därmed kunde genomföra testet var och när de själva önskade. För att möjliggöra kartläggning av försökspersonens beteende med avseende på kontinuitet i genomförandet och antal tolkningsförsök var formuläret utformat på så sätt att varje knapptryck loggades med aktuell tidpunkt.

Resultaten från pilotstudien visade att orden från meningarna i klass 1 tolkades korrekt i 98 % av fallen. Motsvarande resultat från klass 2 och 3 var 88 % respektive 80 % av testfallen. Det förekom relativt stora variationer mellan de olika testdeltagarna. Bland annat uppvisade test genomförda med enkla datorhögtalare som ljudkälla lägre resultat än de där bättre högtalare eller hörlurar använts. Dessutom visade det sig att personer med annat modersmål än svenska lyssnade fler gånger på varje ljud än genomsnittet. Ytterligare en iakttagelse som gjordes utifrån resultaten var att personer som genomförde hela testet utan avbrott lyckades bättre än testdeltagare som gjorde ett eller flera uppehåll²⁰. I flertalet fall var minst två av dessa omständigheter kombinerade på olika sätt.

7.2 Laborietest med försökspersoner

Den kunskap som förvärvats från analysen av pilotstudien användes för att utforma ett andra test, med syfte att erhålla underlag för prövning av hypoteserna som ställts i detta arbete. För att undvika resultatvariationer orsakade av hårdvaruskillnader och externa störningsmoment bestämdes att testet skulle genomföras i laboriemiljö.

7.2.1 Utformande av testmängd

Hypotes 1 innebär ett antagande om att DAG kan representera ljud med en högre densitet än vad som kan uppnås med digitala representationstekniker. I kapitel 6

²⁰ En uppföljning av mycket avvikande resultat gav vid handen att några av testerna hade genomförts parallellt med arbete, telefonsamtal och dylikt, vilket inte bara förklarade diskontinuiteten utan även antydde att uppgiften krävde en viss grad av koncentration.

uppvisades den högsta densiteten, 390 bytes per cm^2 eller 2,6 cm^2 per ljudsekund, av en DataMatrix-symbol som var tolkningsbar i 47 % av testfallen. Om en DAG med samma eller högre densitet uppvisade en högre tolkningsbarhet skulle detta vara till hypotesens fördel, om tolkningsbarheten var lägre vore detta till hypotesens nackdel. Därför skapades en klass DAG-representationer med en densitet av 2,4 cm^2 per ljudsekund. Denna klass kommer att refereras till som klass A.

För att ge mera underlag för en hypotesprövning undersöktes även tolkningsbarheten av DAG-representationer med högre densitet än 2,4 cm^2 per ljudsekund. Eftersom pilotstudien visat att representationer med dubbelt så hög densitet var tolkningsbara i viss utsträckning skapades även en klass med denna densitet, det vill säga 1,2 cm^2 per ljudsekund, som kommer att refereras till som klass B. En tredje klass, C, med ytterligare en fördubbling av densiteten, 0,6 cm^2 per ljudsekund, skapades med syftet att "bottna" undersökningen, det vill säga att hitta en nivå där tolkningsbarheten närmar sig noll.

Liksom i pilotstudien varierades testmängden med avseende på olika typer av röster. Tre röster med skilda karaktäristika var representerade: en mörkare mansröst (Tomas Bolme), en något ljusare mansröst (Fredrik Lindström) samt en kvinnorröst (Katarina Ewerlöf). Urvalet av meningar skedde med målet att försöka minimera risken för kontextbaserade tolkningar, det vill säga att ledtrådar till vilka ord som sägs kan erhållas genom att lyssnaren identifierar röstens ägare. Exempelvis undveks meningar som innehöll ord som *riddare* eller *rustning*, eftersom en lyssnare som vet att Tomas Bolme är uppläsare av böcker med detta tema kan få vägledning i tolkningsprocessen genom att känna igen rösten. Var och en av rösterna förekom en gång per densitetsklass, vilket resulterade i att totalt nio olika meningar ingick i testmängden.

Hypotes 2 innebär att DAG är mer tolerant gällande hårdvaruvariationer än vad digitala representationer är. För att ge underlag för prövning av denna hypotes utformades testmängden i två olika varianter. I den ena var vartannat testfall resultatet av en DAG inskannad med en maximal skannerupplösning, medan övriga var inskannade med hälften så hög skannerupplösning. Den andra varianten skapades på omvänt vis. Blandningen gjordes för att underlätta beslut om huruvida enskilda resultat beror på skannerupplösning eller på den aktuella personens perceptionsförmåga. På grund av resursbegränsningar väntades antalet testdeltagare inte bli tillräckligt högt för att testa variationer med avseende på skrivrupplösning. Just denna variation uteslöts med bakgrund av att den i kapitel 4 bestämdes ha lägst prioritet.

Samtliga 18 testljud finns online på <http://www.his.se/iki/dagdemo>

7.2.2 Genomförande av test

Interaktiva testformulär liknande det som använts i pilotstudien skapades för de båda varianterna av testmängden²¹. Testljuden grupperades klassvis så att de tre meningarna från klass A var först i testformuläret, därefter meningarna i klass B och sist de tre meningarna från klass C. På så sätt eskalerede densiteten hos representationerna och därmed den förmodade svårighetsgraden för tolkning av innehållet.

²¹ De båda varianterna alternerades så att varannan försöksperson testade den ena varianten och varannan den andra.

Som nämndes inledningsvis genomfördes testet i ett laboratorium, vilket innebar att alla tester gjordes under identiska förhållanden med avseende på miljö och hårdvara. Bortsett från relativt högljudd utrustning var laboratoriet isolerat från störningsmoment som kunde distrahera testdeltagaren. På grund av utrustningens ljudnivå användes hörlurar anslutna till en förstärkare, på vilken försökspersonen själv enkelt kunde justera volymen till önskad nivå.

Totalt deltog 16 personer i experimentet, 8 kvinnor och 8 män. Samtliga 16 var anställda eller studerande vid högskolan i Skövde, och tillfrågades antingen muntligt eller via anslag om de ville delta i testet. Med bakgrund av resultaten från pilotstudien beslöts att endast inkludera personer med svenska som modersmål. Varje försöksperson informerades både muntligt och skriftligt om sin uppgift: att försöka höra vad som sades i de nio meningarna samt skriva de ord som uppfattades. Vidare informerades om att testet inte var begränsat med avseende på tid eller antal försök att tolka ett ljud.

7.2.3 Resultat

Försökspersonernas svar utvärderades enligt två olika modeller. Den ena modellen innebar en beräkning av andelen korrekt tolkade ord i varje mening. Därefter beräknades den genomsnittliga andelen korrekta ord för varje densitetsklass. Beräkningarna gjordes dels totalt över klassen, dels separerat med avseende på vilken av de två skannerupplösningarna som använts vid inläsning av aktuell representation. Detta sätt att räkna mäter således graden av informationsmässig ekvivalens mellan två representationer, där de atomära delarna utgörs av ord. I den andra modellen för att bedöma informationsmässig ekvivalens betraktades meningarna som atomära. Detta kan liknas vid sättet som digitala representationer tolkas på: antingen tolkas all information, eller så tolkas ingen information. Således ger en helt korrekt tolkad mening ett positivt utfall och alla andra tolkningar negativa utfall. Resultatet från de båda olika beräkningssätten presenteras nedan (se även bild 33 och bild 34).

Samtliga meningar med densitetsklass A, det vill säga representationer med densiteten $2,4 \text{ cm}^2$ per ljudsekund uppfattades helt korrekt av samtliga deltagare, oavsett vilken skannerupplösning som använts.

Av meningarna från densitetsklass B, det vill säga representationer med densiteten $1,2 \text{ cm}^2$ per ljudsekund, var i genomsnitt 76 % av meningarna och 89 % av orden korrekt tolkade då en hög skannerupplösning använts. Från de representationer som var skannade med låg skannerupplösning tolkades 57 % av meningarna och 82 % av orden korrekt. Var och en av de tre meningarna tolkades helt korrekt av minst två försökspersoner för respektive skannerupplösning.

I densitetsklass C, det vill säga representationer med densiteten $0,6 \text{ cm}^2$ per ljudsekund tolkades 9 % av meningarna och 50 % av orden korrekt vid hög skannerupplösning. Låg skannerupplösning resulterade i att 4 % av meningarna respektive 29 % av orden tolkades korrekt. Endast tre meningar tolkades helt korrekt i denna klass, och det bör nämnas att det var en och samma person som lyckades tolka alla tre.

7 Densitetstest av DAG

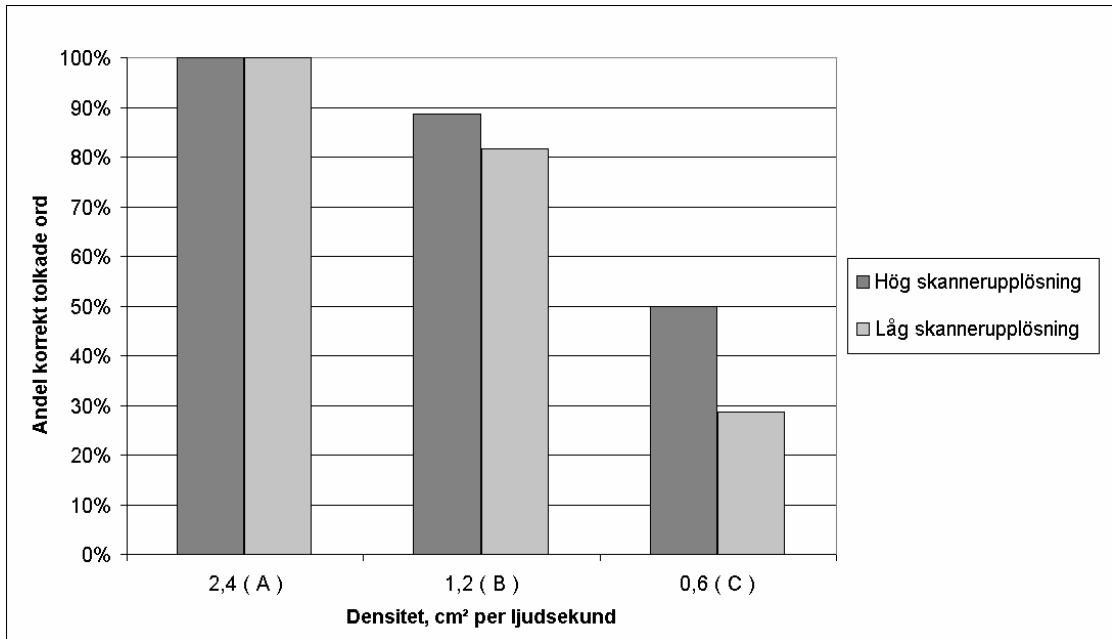


Bild 33: Genomsnittlig andel korrekt tolkade ord för de tre densitetsklasser som ingick i användartestet. Mörka stolpar gäller representationer som är skannade med hög upplösning (600 dpi), medan ljusa stolpar visar resultatet från en låg skannerupplösning (300 dpi).

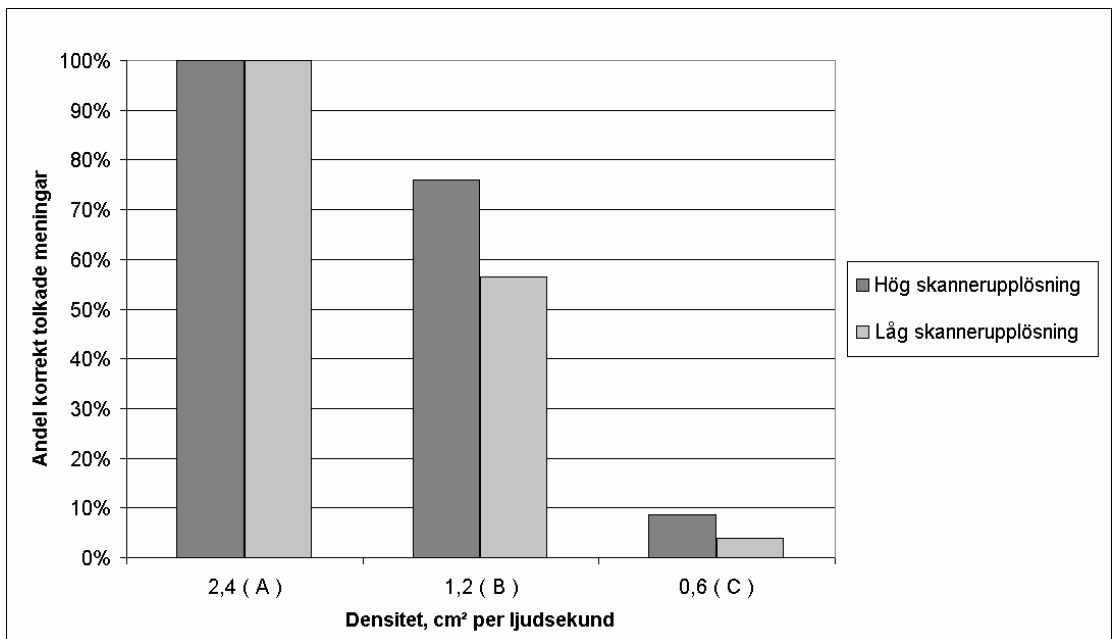


Bild 34: Genomsnittlig andel helt korrekt tolkade meningar för de tre densitetsklasser som ingick i användartestet. Mörka stolpar gäller representationer som är skannade med hög upplösning (600 dpi), medan ljusa stolpar visar resultatet från en låg skannerupplösning (300 dpi).

7.2.4 Kompletterande test av DAG med låg skrivarupplösning

Det låga antalet försöksdeltagare medförde att det inte gjordes några laborietest för att undersöka hur låg skrivarupplösning påverkar tolkningsbarheten av DAG. Därför utfördes ett kompletterande experiment som behandlar detta, där fyra personer som inte ingått i någon av de tidigare testgrupperna deltog. De tre representationer som utgjorde klass A i ovanstående experiment skrevs ut med låg upplösning och

skannades in med hög upplösning. De genererade ljudfilerna skickades till testdeltagarna via e-post, och mottagarna informerades om sin uppgift: att lyssna på ljuden och skriva vad som sades. Resultaten visade att endast en av meningarna tolkades korrekt, vilket gjordes av tre personer. Eftersom inte samtliga meningar var tolkningsbara är värdet av statistiska beräkningar lågt eller rent missvisande här.

7.3 Testanalys

Resultaten visar generellt att tolkningsbarheten av DAG dels är avtagande med ökad densitet och dels är beroende av lyssnarens perceptuella förmåga. I klass A, som hade högst densitet, tolkades samtliga meningar av samtliga testdeltagare. Ljuden från klass B, som hade dubbelt så hög densitet, tolkades korrekt av färre testdeltagare. Majoriteten av tolkningarna från klass B var dock korrekta, både med avseende på andelen korrekta ord och andelen korrekta meningar. En mer dramatisk försämring av tolkningsbarheten uppvisades av ljuden i klass C, som hade en fyra gånger så hög densitet som klass A. Här kunde endast en minoritet av orden och meningarna tolkas korrekt. Dock lyckades en person tolka samtliga tre meningar från klassen, vilket ytterligare understryker att tolkningen i hög grad är beroende av lyssnarens förmågor. Värt att notera är att syftet med klass C var att ”bottna” undersökningen, men att alla meningar i klassen var tolkningsbara av minst en testdeltagare.

Spontana kommentarer från vissa försökspersoner gjorde gällande att de upplevde distorsionen i ljuden som mycket störande och därför försvårade tolkningen. Två personer sade sig å andra sidan vara tränade att höra röster genom brus, eftersom de tidigare hade arbetat med radiokommunikation. Därför upplevde dessa personer inte distorsionen som fullt så störande. Det låga antalet testdeltagare gör det dock svårt att dra några slutsatser om huruvida en sådan vana innebär en markant skillnad för resultaten.

Försämringen av tolkningsbarheten på grund av minskad skannerupplösning tilltog exponentiellt med ökad densitet. För klass A hade skannerupplösningen ingen märkbar påverkan av tolkningsbarheten. I klass B orsakade minskad skannerupplösning en försämrad tolkningsbarhet av meningar med cirka 25 %, medan tolkningsbarheten av ord sänktes med cirka 10 %. För klass C halverades tolkningsbarheten både med avseende på hela meningar och ord.

Minskad skannerupplösning får således samma effekter som ökad densitet, det vill säga att en inskannad bild blir mer grovmaskig. Därmed blir också approximationerna av inläst data mindre noggranna, vilket får till följd att skillnaderna mellan ljudet från representationen och det representerade ljudet ökar. Av samma orsak resulterar också en minskad skriverupplösning i sämre tolkningsbarhet, om än mer drastiskt. Testerna visar dock att hårdvaruvariationer såväl som densitetsvariationer inte medför några skarpa gränser för tolkningsbarheten, utan istället leder till successivt avtagande respektive tilltagande kvalitet beroende på vilka variationer som görs.

8 Jämförelse mellan DAG och digitala streckkoder

I detta kapitel jämförs egenskaperna hos DAG med de digitala streckkoderna utifrån resultaten från de underökningar som gjorts i kapitel 6 och 7. Densiteten hos representationerna beskrivs i termer av cm^2 per ljudsekund, där ljudet för de digitala streckkoderna är maximalt komprimerat med mp3. Tolkningsbarheten för ljud med denna kompressionsgrad är inte undersökt men antas vara 100 %. Dessutom antas de digitala representationerna ha en obegränsad lagringskapacitet, vilket inte är fallet i praktiken (se kapitel 6). Dessa antaganden innebär att de faktiska densitetvärdena för digitala streckkoder troligtvis är något lägre än de som används vid jämförelserna nedan.

De fundamentala skillnaderna mellan teknikerna som jämförs bör för tydlighets skull belysas ännu en gång. Jämförelserna nedan gäller informationsbevarande egenskaper med avseende på talade budskap, trots att de digitala teknikerna i motsats till DAG inte är begränsade till representation av ljud. Dessutom resulterar en lyckad tolkning av digital information i en kopia av ursprunglig data, oavsett vad som är representerat. En tolkning av DAG innebär däremot aldrig fullständigt återskapande av representerad data, utan i bästa fall en tolkningsbar approximation av densamma²².

8.1 Hög skannerupplösning

DAG-ljud visade sig vara tolkningsbara i alla eller nära alla testfall upp till och med en densitet av $2,4 \text{ cm}^2$ per ljudsekund. DataMatrix-symboler med samma densitet var tolkningsbara i 47 % av testfallen. Detta var också den högsta densitet som uppmättes för digitala streckkoder. Först vid 4 cm^2 per ljudsekund gav alla eller nära alla DataMatrix-symboler positiva testresultat, vilket innebär att densiteten hos DAG-representationer med samma tolkningsbarhet var drygt 1,5 gånger så hög. Ytterligare en dubbling av densiteten för DAG till $1,2 \text{ cm}^2$ per ljudsekund resulterade i att 76 % av representationerna tolkades korrekt. Detta är fortfarande ett högre värde än vad som uppmättes för en DataMatrix-symbol med mindre än hälften så hög densitet. Andelen korrekt tolkade ord var för DAG vid denna densitet 82 %. Ännu en dubbling av densiteten resulterade i att så gott som samtliga tolkningsförsök av hela meningar underkändes, medan hälften av orden tolkades korrekt (se bild 35).

²² Detta gäller indirekt tolkning, det vill säga representationer som är utskrivna och inlästa, alternativt manipulerade med digital bildbehandling såsom förlustgenererande komprimering eller dylikt. Direkt tolkning av DAG kan däremot göras helt förlustfritt.

8 Jämförelse mellan DAG och digitala streckkoder

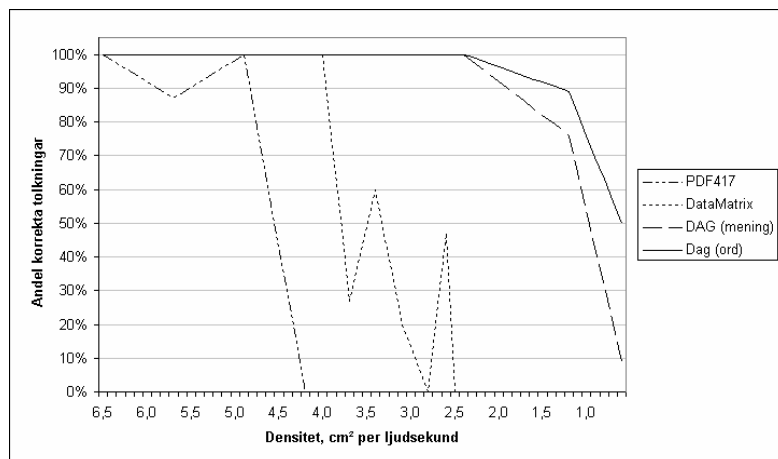


Bild 35: Jämförelse av andelen korrekta tolkningar för respektive representationsform vid hög skannerupplösning (600 dpi).

8.2 Låg skannerupplösning

Vid en halvering av skannerupplösningen visade sig DAG-ljud fortfarande vara tolkningsbara i alla eller nära alla testfall upp till och med en densitet av 2,4 cm² per ljudsekund. Den högsta uppmätta densiteten för digitala streckkoder återfanns hos DataMatrix-symboler med densiteten 6,6 cm² per ljudsekund. Tolkningsbarheten hos dessa symboler var 60 %, vilket endast är tre procentenheter högre än tolkningsbarheten hos DAG-representationer med drygt fem gånger så hög densitet (1,2 cm² per ljudsekund). Den högsta densiteten för digitala streckkoder som var tolkningsbara i alla eller nära alla testfall var 10,3 cm² per ljudsekund. DAG-ljud med motsvarande tolkningsbarhet hade drygt fyra gånger så hög densitet (se bild 36).

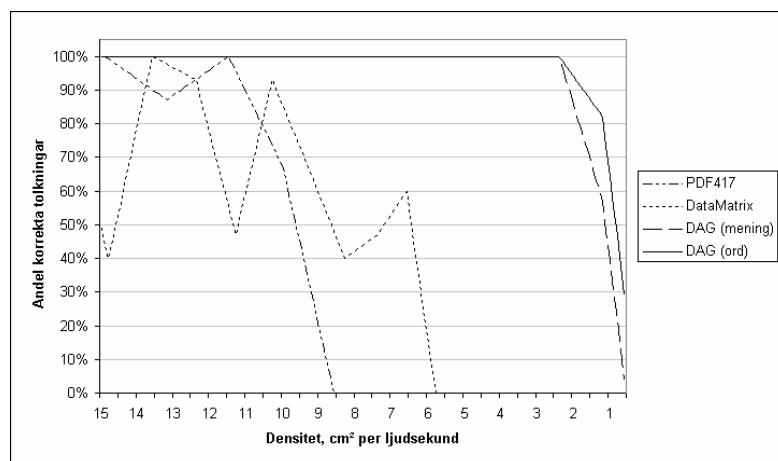


Bild 36: Jämförelse av andelen korrekta tolkningar för respektive representationsform vid låg skannerupplösning (300 dpi).

9 Slutsatser

Ett mål med detta arbete var att skapa en teknik för att illustrera hur ett ljuds analoga egenskaper kan utnyttjas i grafiska representationer. Bristen på existerande tekniker av detta slag gjorde det svårt att på förhand veta något om dess praktiska egenskaper, även om de teoretiska förutsättningarna var lovande (se kapitel 2 och kapitel 3). Utvecklingens initiala faser, det vill säga arbetet som föregick det första uppspelade ljudet, präglades därför av stor ovisshet. Resultatet var emellertid överraskande lyckat, och uppmuntrade till en vidareutveckling som var mer omfattande än vad som hade planerats från början. Tekniken kom så småningom att utvecklas till ett patentsökt²³ koncept för lagring och överföring av ljudinformation – DAG. Trots prototypstatus har DAG visat prov på hög lagringskapacitet och god flexibilitet, parametrar som är avgörande för prövning av de båda hypoteser som ställts i arbetet.

I detta kapitel redovisas hypotesprövningarna följt av en diskussion om möjliga applikationsområden för DAG. Därefter utvärderas relevansen hos de resultat som ligger till grund för hypotesprövningarna. Avslutningsvis föreslås ett antal tänkbara framtida arbeten, som ur olika perspektiv är relaterade till de genomförda undersökningarna.

9.1 Prövning av hypotes 1

Den första hypotesen som ställdes innebar att en analog grafisk ljudrepresentation kunde ha en högre densitet än en digital dito. För att möjliggöra en jämförelse antogs två ljudrepresentationer vara informationsmässigt ekvivalenta om båda kunde förmedla innehållet i ett talat budskap.

De streckkoder som hade högst densitet kunde tolkas av den använda mjukvaran i ungefär hälften av de försök som gjordes. Samtliga ljud som representerats av DAG med denna densitet kunde tolkas av alla 16 lyssnare vid ett test i laboratorium. Samma test visade att DAG-ljud med ytterligare dubbelt så hög densitet var tolkningsbara i ungefär tre fjärdedelar av testfallen. Utifrån detta perspektiv anses därför hypotes 1 vara styrkt.

9.2 Prövning av hypotes 2

Den andra hypotesen gjorde gällande att en analog grafisk ljudrepresentation var mer tolerant för hårdvaruvariationer än en digital dito.

De minsta digitala streckkoderna som var tolkningsbara i alla eller nära alla testfall med hög skannerupplösning gick inte att tolka då skannerupplösningen halverades. För tolkningsbarheten av DAG innebar en sänkt skannerupplösning ingen skillnad för representationer som tidigare var tolkningsbara i alla eller nära alla testfall. För de representationer som med hög skannerupplösning var tolkningsbara i tre fjärdedelar av testfallen reducerades tolkningsbarheten med ungefär 15%. Därmed anses hypotes 2 vara styrkt med avseende på variationer i skannerupplösning.

En sänkt skrivarupplösning visade sig påverka tolkningsbarheten hos samtliga representationsformer negativt. Densitetssänkningen för de digitala streckkoderna motsvarade ungefär den försämring som uppmättes vid en sänkning av

²³ DAG är registrerat som *Patent Pending* vid United States Patent and Trademark Office.

skannerupplösningen. För DAG orsakade den sänkta skriverupplösningen en mer dramatisk försämring än vad en sänkt skannerupplösning gjorde. På grund av ett lågt antal testdeltagare ingick denna variation inte i testerna av DAG, utan genomfördes som ett enklare kompletterande test. Därför har en noggrann kartläggning av försämringen inte kunnat genomföras, men det kan fastställas att DAG inte är okänslig för skriverupplösning i den utsträckning som krävs för att styrka hypotes 2. Denna hypotes måste därför revideras till att enbart gälla variationer av inläsningshårdvara.

9.3 Användningsområden för DAG

Utöver långtidsarkivering av ljud, som beskrevs i kapitel 3, har ett flertal tänkbara tillämpningar av DAG identifierats under arbetets gång. Ett urval av dessa tillämpningar presenteras nedan.

9.3.1 Verktyg för upphovsrättsskyddande av ljud

Vissa ljud används som varumärken, exempelvis Hemglass-bilens melodi och filmbolaget MGM:s rytande lejon. För att upphovsrättsskydda ett varumärke, oavsett om det är en logotyp eller ett ljud, måste det kunna representeras grafiskt (Löfgren, 2003). Om varumärket består av en melodi kan den grafiska representationen utgöras av notskrift, medan andra ljud (som exempelvis ett rytande lejon) är hänvisade till att representeras med sonogram (Löfgren, 2003). Ett sonogram är en tvådimensionell kurva som beskriver ett ljud som amplitudförändringar över tid. Amplituden utgör y-axeln i kurvan, vilket innebär att ett sonogram är mer utrymmeskrävande än en DAG, eftersom DAG beskriver amplituder med färgnyanser. Det är således möjligt att med DAG grafiskt beskriva ett ljud på en yta som är mindre än ytan som upptas av ett motsvarande sonogram.

9.3.2 Artefakter

DAG kan användas för att enkelt och till en låg kostnad materialisera ljud, det vill säga skapa exempelvis samlarkort, flygblad eller direktreklam. På samlarkort för barn, exempelvis Pokémon-kort, skulle en DAG som beskriver den aktuella figurens läte kunna tryckas. Med tanke på de toleranta hårdvarukraven vid tolkning skulle en enkel skanner för ändamålet kunna tillverkas för ett överkomligt pris. Tidskrifter kan locka köpare genom att erbjuda smakprov på ny musik, tryckt på samma papper som resten av innehållet. Genom att ljudkvaliteten kan begränsas innebär detta ingen konkurrens för skivförsäljning, tvärtom kan metoden användas av musikindustrin för marknadsföring.

9.3.3 Distribuerade kvalitetsprov för skrivare

Resultaten visade att kvaliteten hos en DAG är mer beroende av utskriftskvalitet än av skannerupplösning. Detta faktum kan göra DAG till ett användbart verktyg vid utvärdering av skrivare. Vid exempelvis online-handel är det svårt för en kund att jämföra utskriftskvaliteten mellan två skrivare på annat sätt än att jämföra tekniska specifikationer. Som ett komplement skulle det vara möjligt att online jämföra två DAG-ljud utskrivna med de aktuella skrivarna, eftersom ljud är mer lätt distribuerat än faktiska utskrifter. Det ljud som har högst kvalitet är sannolikt utskrivet med den skrivare som har bäst återgivning.

9.4 Resultatens relevans

Undersökningen av de digitala streckkoderna innebar ett antal antaganden som å ena sidan var nödvändiga för undersökningen, men å andra sidan utgör en grund för ifrågasättande av korrektheten hos resultaten. Experimenten utfördes med generell kontorsdatorutrustning, vilket innebär att ingen hänsyn togs till huruvida teknikerna kan uppvisa bättre prestanda med hårdvara som är speciellt avsedd för respektive teknik. Dessutom pekade en oregelbundenhet i testresultaten på att vissa symbolstorlekar var mindre lämpade för aktuella hårdvaruinställningar än andra. På grund av tid- och utrymmesbegränsningar genomfördes ingen närmare analys av detta fenomen. I undersökningen testades dock ett relativt stort antal symbolstorlekar med relativt små storleksvariationer sinsemellan. Symbolerna skrevs ut i flera exemplar, och på varje utskrift utfördes flera inskanningar och tolkningsförsök. De minsta tolkningsbara symbolerna var dessutom av mellanstorlek, det vill säga att det för respektive teknik fanns flera symbolstorlekar som var mindre än de minsta tolkningsbara symbolerna. Resultaten kan därför anses vara giltiga för den typ av utrustning som användes vid experimenten. Däremot kan ingenting sägas om resultatens giltighet för hårdvara som är speciellt avsedd för digitala streckkoder.

Då de digitala streckkodernas lagringskapacitet skulle beräknas i termer av tid användes maximalt komprimerat mp3-ljud som teoretisk måttstock. Valet av mp3 baserade sig på formatets utbredda användande snarare än en uttömmande undersökning av olika komprimeringsmetoder. Frågan om det finns metoder som erbjuder än mer effektiv komprimering av generella ljud måste därför lämnas öppen. En metod som är marginellt mer effektiv än mp3 utgör dock inte ensam något hot mot de övergripande resultaten i detta arbete.

Angående testerna av DAG medför det låga antalet testdeltagare tillsammans med det begränsade antalet testljud att resultaten får ses som ungefärliga. Detta gäller i synnerhet klassen med högst densitet där den uppmätta tolkningsbarheten var låg. I klassen med lägst densitet var däremot andelen lyckade tolkningar 100 %, vilket är högt nog för att de uppmätta resultaten kan anses vara tillförlitliga. Det bör understrykas att testerna inte var avsedda att mäta några andra kvalitéer hos ett ljud än just dess duglighet att förmedla talade budskap. Detta trots att varken DAG eller mp3 är begränsade till denna typ av ljud.

9.5 Framtida arbeten

Arbetet har spänt över flera olika områden samtidigt som tid och utrymme har varit begränsat. Detta har medfört att de olika undersökningarna sällan varit mer djupgående än vad som har krävts för att uppnå tillförlitliga resultat. På grund av detta lämnas en rad frågor öppna för framtida arbeten inom de olika områdena.

9.5.1 Fortsatt utveckling av DAG

En vidareutveckling av DAG kommer först och främst att fokusera på ökad läsbarhet, dels för att uppnå än högre densitet, dels för att underlätta inläsning med exempelvis digitalkamera. En åtgärd för att ytterligare förenkla tolkningsprocessen är att i anslutning till en DAG ange korrekt läshastighet. På grund av den lilla datamängden och vikten av hög precision kan läshastigheten med fördel representeras med en digital streckkod.

Största delen av detta arbete har ägnats åt att undersöka kvantitativa aspekter, det vill säga hur stor yta som krävs för att representera ett sekundlångt ljud. För att öka användbarheten av DAG bör i framtiden även kvalitativa aspekter behandlas, vilket kan komma att kräva justeringar av algoritmer för såväl skapande som inläsning av DAG.

Det finns dessutom planer på att utveckla en variant av DAG som använder två färger istället för de gråskalor som används i nuvarande version. En introduktion av färger skulle eventuellt göra DAG mer estetiskt tilltalande, men framför allt öka tillämpbarheten vid kvalitetstester av skrivare som beskrevs i kapitel 9.3.3.

9.5.2 Kartläggning av kognitiva parametrar

Resultaten från tester med försökspersoner visade prov på komplexiteten hos den mänskliga perceptionen. Det statistiska underlaget var dock inte tillräckligt stort för att kartlägga vilka parametrar som påverkar tolkningen av talat språk. Exempel på parametrar som skulle kunna vara av betydelse är uppläsarens röstläge, uppläsarens talhastighet och lyssnarens vana att uppfatta ord genom brus. Tänkbara framtida arbeten skulle kunna behandla ämnen som:

- Orsaker till att vissa personer lyckas tolka ljudinformation bättre än andra, trots fysiskt lika förutsättningar.
- Skillnader mellan hur olika typer av förvrängningar av ljud påverkar tolkningsbarheten. Med förvrängningar menas reducering av data, tillägg av brus eller en kombination av de båda.
- Kontextens betydelse för tolkning av språk under brusiga förhållanden, alltså på vilket sätt tolkning av enstaka ord kan underlättas genom att infoga dem i längre meningar.

Ytterligare en frågeställning är i vilken utsträckning olika former för grafisk ljudrepresentation kan ge visuella ledtrådar om hur ett ljud låter. Det vill säga huruvida det är möjligt att få en uppfattning om ett ljud endast genom att se på notskrift, sonogram eller DAG.

9.5.3 Utforskning av digitala tekniker

Ett mål med utvecklingen av DAG har varit att påvisa teknikens flexibilitet och lagringskapacitet i praktiken jämfört med digitala tekniker. Utgångspunkten för jämförelserna har varit PDF417 respektive DataMatrix kombinerat med mp3, med andra ord tekniker som är standardiserade och allmänt utbredda. En tänkbar uppgift är att genomföra en undersökning med motsatt mål, det vill säga visa en digital teknik med bättre egenskaper än DAG. En sådan teknik skulle, givet liknande förutsättningar med avseende på hårdvara och jämförelsevillkor, vederlägga hypoteserna i detta arbete.

Referenser

Avison, D.E. och Fitzgerald, G., 1995. *Information systems development: Methodologies, techniques and tools* (2nd ed.), Blackwell, Oxford, Great Britain.

Bresenham, J.E., 1965. *Algorithm for computer control of digital plotter*. IBM Systems Journal., vol. 4, nr. 1, 1965, sid 25-30.

Burke, M. W., 1996. *Image Aquisition*. Chapman & Hall, Oxford, Great Britain.

Carr, R. E. och Hayes, R. M., 1988. *Wide Screen Movies*. McFarland & Co. Inc., Jefferson, North Carolina, USA.

Cost, F., 1997. *Digital Printing*. Delmar Publishers, New York, USA.

ISO/IEC International Standard IS 11172-3. *Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbits/s - Part 3: Audio*

ISO/IEC International Standard IS 13818-3. *Information technology - Generic coding of moving pictures and associated audio information - Part 3: Audio*

ISO/IEC International Standard IS 15438. *Information technology - Automatic identification and data capture techniques - Bar code symbology specifications - PDF417*

ISO/IEC International Standard IS 16022. *Information technology - International symbology specification - Data matrix*

Kantner, L., 1994. *Techniques for Managing a Usability Test*. IEEE Transactions on Professional Communication, Volume 37, Number 3, September 1994.

Löfgren, C., 2003. *Så ser ljudet ut*. Brandnews, nr 10, 2003, sid 8 – 13.

Ohlsson, S., Webb, J. S. och Westerlund, B., 1999. *Digital Bild – kreativt bildskapande med dator*. Bonnier, Stockholm.

Palmer, S. E., 1978. *Fundamental Aspects of Cognitive Representation*. Cognition and Categorization, Rosch, E. and Lloyd, B. B., (eds.), LEA, Hillsdale, NJ, sid 259 - 303.

Pavlidis, T., 2000. *A New Paper/Computer Interface: Two-Dimensional Symbologies*. Proceedings of the International Conference on Pattern Recognition, 2000.

Pavlidis, T., Swartz, J. och Wang, Y. P., 1990. *Fundamentals of bar code information theory*. IEEE Computer, 23(4), April, sid. 74-86.

Priddy, D. G. och Cymbalski R. S., 1990. *Dynamically variable machine readable binary code and method for reading and producing thereof*. U.S. Patent 4 939 354, Juli 1990.

Referenser

Rumsey, F. och McCormick, T., 1994. *Sound & Recording: An Introduction*. Second edition, Focal Press, Great Britain.

Roland-Mieszkowski, M. och Young, W. R., 1991. *Consequences of Nyquist Theorem for Acoustic Signals Stored in Digital Format*. Proceedings from Acoustic Week in Canada 1991 - CAA Conference, Edmonton, Alberta, Canada, October 7 - 10, 1991

Rouse, J. H., 1930. *Photoelectric Cell*. U.S. Patent 1 921 483, November 1930.

Stenning, K. och Oberlander, J., 1995. *A cognitive theory of graphical and linguistic reasoning: logic and implementation*. Cognitive Science, 19, sid 97-140

Sandoval, I. och Lane, T. (contacts), 1999. *Information Architecture White Paper IA-6801: Electronic Image Formats and Compression Algorithms*. Los Alamos National Laboratory Information Architecture Standard, Revision 2.0, Oct 20, 1999.

Smith, A., 1999. *Preservation in the Digital Age: What is to Be Done?* American Libraries, Vol 30, Issue 3. sid 36-39.

Story, M., 1997. *A Suggested Explanation for (Some of) the Audible Differences Between High Sample Rate and Conventional Sample Rate Audio Material*. White Paper, dCS Ltd, UK, September 1997.

Wang, Y. P., 1993. *System for encoding and decoding data in machine readable graphic form*. U.S. Patent 5 243 655, September 1993.