

Formulering av algoritm för prosodimodellen FK-systemet

Theresa Andersson
theresa@stp.ling.uu.se

Examensarbete i datorlingvistik
Språkteknologiprogrammet
Uppsala universitet – Institutionen för lingvistik och filologi

10 september 2004

Handledare:
Fredrik Larsson, Phoneticom AB
Bertil Lyberg, Uppsala universitet

Sammanfattning

Ett flertal prosodimodeller för svenskt tal som har utvecklats med åren bygger på Bruces (1977) uppmärkning av accent 1 och accent 2. Likaså gör den prosodimodell som utvecklats av Gunnar Fant och Anita Kruckenberg, FK-systemet, vilken har analyserats i det här arbetet. Utifrån en text beräknar modellen fonemljuddens duration samt grundtonsfrekvens och genererar syntetiskt tal. Talet har svensk intonation med stockholmsdialekt. Syftet med det här arbetet var att strukturera FK-systemet i en implementerbar form. Implementeringen anpassades till de förutsättningar som systemet krävde samt med tanke på att tillägg och förändringar av regelverket smidigt ska kunna infogas, eftersom modellen utvecklas och förfinas kontinuerligt. För att tydliggöra arkitekturen och beräkningarna som utfördes på en text har detta illustrerats med ett pedagogiskt exempel.

Det vore önskvärt att systemet i framtiden kommer att täcka samtliga centrala delar i den svenska prosodin och därmed även terminal junktur och emfas. Vidare ska systemet kunna komma att användas som den lingvistiska komponenten i ett text-till-talsystem.

Förord

I den här uppsatsen redovisas ett examensarbete som utförts som en del av Språkteknologiprogrammet vid Institutionen för lingvistik och filologi, Uppsala universitet. Examensarbetet har utförts på Phonicom AB i Uppsala.

Jag vill här ta tillfället i akt att tacka de personer som har hjälpt mig att genomföra arbetet. Först vill jag rikta ett varmt tack till Gunnar Fant och Anita Kruckenberg som grundare av FK-systemet har gett mig djupgående instruktioner och klarhet i modellens uppbyggnad i tid och otid. Ett tack går även till min handledare på Institutionen för lingvistik och filologi, Bertil Lyberg, som har kommit med värdefull kritik och goda råd kring uppsatsens innehåll samt för hans uppmuntran. Ytterligare ett tack går till min handledare Fredrik Larsson på Phonicom AB, som gett mig vägledning och konstruktiva idéer i utvecklingen av arbetet.

Ett varmt tack riktas till er alla som funnits i min närhet under arbetets gång!

Theresa Andersson
Uppsala september 2004

Innehållsförteckning

1 Inledning	5
1.1 Syfte	5
1.2 Disposition av uppsatsen	5
2 Bakgrund	6
2.1 Text-till-talsystem	6
2.1.1 Språkbehandling	6
2.1.2 Signalbehandling	7
2.2 Grunder i svensk prosodi	9
2.2.1 Duration	9
2.2.2 Accenter och grundtonsfrekvens	10
2.2.3 Betoning	12
2.3 Prominens och gruppering	12
2.4 FK-systemet	13
2.4.1 Prominens	13
2.4.2 Gruppering	14
2.4.3 Junktur och pauser	14
2.4.4 Final förlängning	14
2.4.5 Accentmodulering	15
2.4.6 Beräkning av grundtonsfrekvens	15
2.4.7 Fonemduration	15
3 Genomförande	17
3.1 Textbearbetning	17
3.2 Gruppering	18
3.3 Accentmodulering	21
3.4 Beräkning av duration	22
3.5 Beräkning av grundtonsfrekvens	23
4 Resultat	25
4.1 Textbearbetning	25
4.2 Tilldelning av paus och final förlängning samt skifte av baskurva	25
4.2.1 Val av typ av baskurva	25
4.3 Tilldelning av grundtonsfrekvens och prominens (R_sF_0)	25
4.4 Tilldelning av prominens (R_sD) och beräkning av duration	26
4.4.1 Final förlängning	27
4.4.2 Klusterreduktion	27
4.5 Tilldelning av relativ position för grundtonsfrekvensnivåer	27
4.6 Beräkning av grundtonsfrekvens	27
4.7 Utdata	28
5 Slutsats	31
5.1 Framtida utveckling	31
Bibliografi	33
Appendix	35

1 Inledning

För att talsyntes ska låta naturligt ställs krav på flera av de ingående delarna i talgenereringen, varav prosodin är en mycket viktig del. Den hjälper oss att förstå ords semantiska betydelse och vad som framhävs i ett yttrande. En onaturlig prosodi eller avsaknad av sådan, gör det svårt att uppfatta ett yttrande. Ett flertal beskrivningar av svensk prosodi har utvecklats med åren främst för att kartlägga svensk prosodi, men även för att användas i generering av talsyntes liksom prosodimodellen FK-systemet.

FK-systemet är utvecklat av Gunnar Fant och Anita Kruckenberg under en femtonårsperiod och finns ursprungligen presenterad i form av ett flertal ark i kalkylprogrammet Excel med tillhörande dokumentation. Den syntaktiska analysen är till stor del stommen i systemet som även innehåller markering av svenskans accent 1 och 2, mätvärden för prominens och beräkningar av värden för fonem- och pausduration samt grundtonsfrekvens. Målet är att via en abstrakt autosegmentellt fonologisk nivå realisera ett yttrandes prosodi. Formatet på modellens utdata är anpassat till ett gränssnitt för talsyntes producerat med hjälp av algoritmen MBROLA.

En generellt gällande algoritm ska utvecklas utifrån den befintliga representationen. Algoritmen tillsammans med definierade förutsättningar och förarbetssteg ligger till grund för en utvärdering av modellens potential till implementering. Om ett godkänt resultat uppnås kan modellen kunna komma att vidareutvecklas av företaget Phoneticom AB i Uppsala, vilka är uppdragsgivare till examensarbetet.

1.1 Syfte

Examensarbetet består i att analysera och automatisera FK-systemet som är en datorlingvistisk modell för prosodi i syntetiskt tal för svenska. Syftet är att utifrån analysen och det regelverk som finns formulera en algoritm i implementerbar form. Utformningen av implementeringen ska vara lättillgänglig och i ett senare skede kunna inkorporeras i ett text-till-talsystem.

1.2 Disposition av uppsatsen

Uppsatsen inleds med ett bakgrundskapitel där ett text-till-talsystem samt ett urval av tidigare utvecklade prosodimodeller för svenska beskrivs översiktligt. I samma kapitel beskrivs även grunderna i svensk prosodi och FK-systemets ingående delar. Därefter följer ett kapitel som beskriver implementeringen av modellen och därmed genomförandet av arbetet. För att klargöra hur prosodimodellens samtliga delar hänger samman och hur olika värden beräknas följer ett pedagogiskt exempel i ett specifikt resultatkapitel. Uppsatsen avslutas med ett kapitel som innehåller slutsatser och framtida utveckling.

2 Bakgrund

2.1 Text-till-talsystem

Ett text-till-talsystem genererar syntetiskt tal från en skriven text. Systemet består av en språkbehandlingsdel och en signalbehandlingsdel som i sin tur är indelade i flera moment. Språkbehandlingen består av olika moment av lingvistisk analys som ger upphov till en fonetisk transkription innehållande yttrandets prosodi. I signalbehandlingen genereras syntetiskt tal av transkriptionen och prosodin (Dutoit, 1996).

2.1.1 Språkbehandling

I det första steget i språkbehandlingsdelen går texten igenom en textformatering och de syntaktiska beståndsdelarna identifieras. Den syntaktiska analysen delar in texten i enheter som paragrafer, meningar, fraser och ord. Enheterna analyseras i sin reella kontext för att även möjliggöra identifiering och parsning av ambiguiteter, felstavade ord och icke-grammatiska konstruktioner. För att transformeringen från text till tal ska bli fullständig är det viktigt att hela texten får en syntaktisk tolkning trots eventuella felaktiga konstruktioner (Ceder & Lyberg, 1992:1151f).

Orden tilldelas sedan sina transkriberade motsvarigheter, vilket sker med hjälp av lexikon och transkriberingsregler, sk. letter-to-sound rules. I lexikonet finns de vanligast förekommande orden representerade tillsammans med sin fonetiska motsvarighet. Då ett ord inte hittas i lexikonet sker transkriptionen med transkriberingsreglerna. Reglerna transformerar ortografin till en fonetisk representation. Ofta finns även annan fonetisk information som stavelsegränser och accent eller betoningstyp lagrat för varje lexikoningång. En stavelses betoningstyp informerar om stavelsen är obetonad, primär- eller sekundärbetonad (Frid, 2003:51).

Det sista delmomentet i språkbehandlingen är att modellera de suprasegmentella särdragen i det syntetiska talet. Det finns ett flertal teorier och algoritmer utvecklade för att generera särdragets akustiska korrelat; grundtonsfrekvens, duration och intensitet. De metoder som används i skilda text-till-talsystem för att få ett syntetiskt tal med en naturlig prosodi är främst neurala nätverk, stokastiska eller regelbaserade (Monaghan, 1992).

Prosodimodeller

Prosodimodeller utvecklade för svenskt tal är oftast regelbaserade. Ett flertal modeller har utvecklats med åren och en av de första är en prosodimodell som Carlsson och Granström utvecklade år 1973. Modellens två huvudsakliga komponenter är en linjärt fallande meningsintonation med en konstant start- och ändpunkt samt ordbetoningspunkter som placeras på intonationskurvan. Punkternas placering beror på betoning, accent och segmentets duration. För ord med accent 1 (akut) placeras ordbetoningen på samma nivå som komponenten för meningsintonationen medan den för ord med accent 2 (grav) beräknas utifrån durationen av föregående ord. En cosinuskurva sammankopplar ordbetoningspunkterna och en grundtonsfrekvenskurva genereras för meningen (Carlsson & Granström, 1973).

Den prosodibeskrivningen Bruce utvecklade 1977 har efteråt vidareutvecklats och förfinats och ligger därför som grund i flera senare utvecklade modeller. Modellen består av tre typer av regler: grundtonsfrekvensregler för en relativ nivå av grundtonsfrekvensen, grundtonsfrekvensregler för en faktisk punkt av grundtonsfrekvensen och sammankopplingsregler. Modellen tar en fonetisk transkription av ett yttrande där betoning, accent och junktur är uppmärksatta. Grundtonsfrekvensreglerna för en relativ

grundtonsfrekvensnivå märker upp textens relativa grundtonsfrekvensnivåer utifrån transkriptionen. Nivåerna ger information om dess temporala placering samt relativa grundtonsfrekvens och representeras med H (hög) eller L (Låg). Modellen skiljer mellan ord med accent 1 och accent 2. Skillnaden mellan de två accenterna är den temporala länkningen var nivåerna för den relativa grundtonsfrekvensen placeras i tidsled i yttrandet relativt den betonade stavelsen. H placeras i stavelsen som föregår den betonade stavelsen i ett ord med accent 1, medan L placeras i den primärbetonade stavelsen i ett ord med accent 2. Grundtonsfrekvensreglerna genererar en faktisk grundtonsfrekvens från en relativ grundtonsfrekvensnivå. Inga konkreta frekvensvärden ingår i systemet utan representeras av nivåerna 1-4 där 1 representerar den lägsta nivån. De punkter som ligger på nivå 1 betecknas alltid som L och de på 3 och 4 som H. Både L och H kan vid olika tillfällen förekomma på nivå 2 beroende på om L eller H föregår. För att få en kontinuerlig grundtonsfrekvenskurva av ett yttrande sammankopplas slutligen punkterna med sammankopplingsreglerna genom linjär interpolation (Bruce, 1977:129ff).

Tidigt utvecklade system och algoritmer för svensk prosodi var huvudsakligen utvecklade utifrån tal inspelat i laboratorium. Modellerna som kommit till under de senaste tio åren har utvecklats för att även kunna hantera dialoger och konversationer. Mer lingvistisk information är nödvändig och exempelvis information om lexikal semantik och diskurs är inkluderad (Frid, 2003:78). Den prosodimodellen som Horne och Filipsson utvecklade år 1996 består av minimal syntaktisk information i kombination med lexikal och morfologisk semantik för att generera en prosodisk struktur av svenska texter. Modellen innehåller bland annat en spårningsfunktion för att avgöra om ett innehållsord påträffades för första gången, dvs. *ny*, eller om det stötts på förut, *given*. Initialt tokeniseras texten och sedan slås varje ord upp i ett domänspecifikt lexikon. Spårningsfunktionen tar sedan vid och kontrollerar om varje ord, eller synonym till ordet, stötts på tidigare inom paragrafen och märker ordet med *ny* eller *given*. Orden ordklasstaggas samt klassificeras som antingen funktions- eller innehållsord. Satsgränser identifieras. En prosodisk parser parsar orden i en hierarkisk struktur med tre nivåer: prosodiskt ord, prosodisk fras och prosodiskt yttrande. Ett innehållsord grupperas med funktionsord för att bilda ett prosodiskt ord. En prosodisk fras bildas av en sats som innehåller ett emfatiskt innehållsord. Om ett emfatiskt innehållsord inte ingår i en sats läggs den satsen till närmast föregående prosodiska fras så att de tillsammans bildar en prosodisk fras. Ett paragrafslut avgränsar ett prosodiskt yttrande. Den prosodiskt parsade texten tolkas slutligen och genererar en grafisk representation med trädstruktur (Horne och Filipsson, 1996).

2.1.2 Signalbehandling

Slutligen genereras en talsignal av de lingvistiska data som genererats. De olika modellerna för genereringen av talsignalen skiljer sig åt med avseende på talkvalitet, minnesåtgång, algoritmkomplexitet och hastighet (O'Shaughnessy, 2000:337).

Konkatenerad talsyntes

Den enklaste modellen konkatenerar lagrade segment av förinspelat tal. Det är viktigt att transitionen mellan två ljud fångas för att eventuella koartikulationseffekter ska kunna bibehållas (Bhaskararao, 1994). Segmenten som kopplas samman kan se olika ut. Vanligt är att de lagrade segmenten består av sk. difoner. Gränsen mellan två difoner i stavelsen konsonant-vokal-konsonant ligger mitt i vokalen, där koartikulationen är som minst. Förutom difoner kan större segment som trifoner och halvstavelser konkateneras. Segmenten finns representerade i en databas. Övergångarna mellan de konkatenerade segmenten måste förfinas för att undvika oregelbundenheter i det producerade talet (O'Shaughnessy, 2000:339).

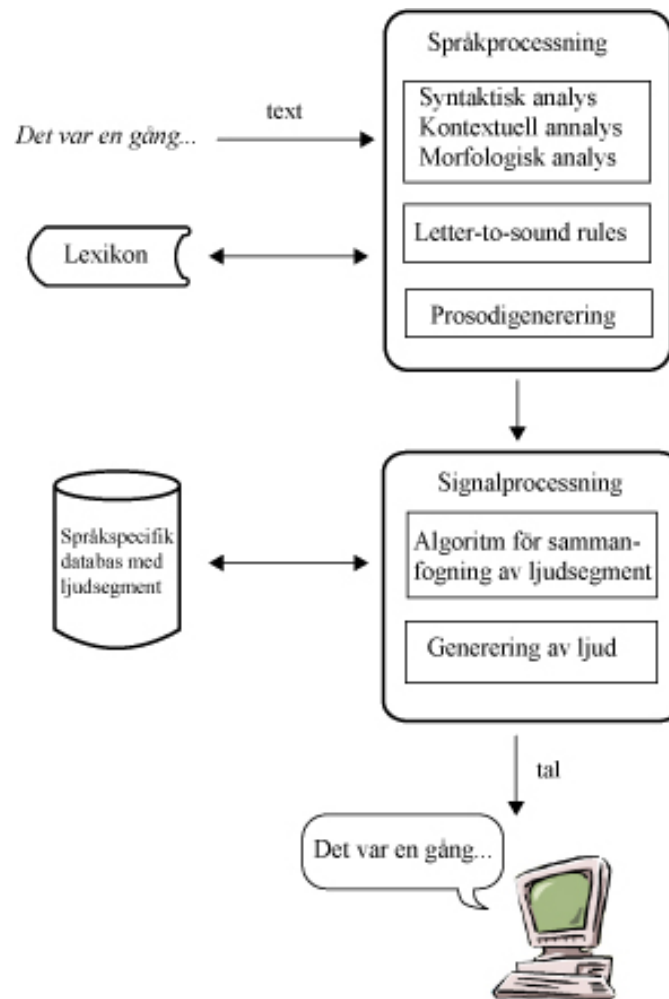
Modellerna för hur signalen fogas samman är många. Den teknik FK-systemet är konstruerat för att använda, MBROLA (Multi Band Resynthesis OverLap Add), är utvecklad på det polytekniska universitetet i Mons i Belgien och är en algoritm som konkatenerar difoner. Difonerna finns representerade i en språkspecifik databas. Som ett första steg i syntesgenereringen med MBROLA läses en textfil in. I textfilen består varje rad av en transkriberad allofon associerad med dess duration i millisekunder. Efter durationen följer en procentsats av durationen för var i allofonen den angivna grundtonsfrekvensen ska uppnås. Ett eller flera värden för procent samt för grundtonsfrekvens i Hertz kan följa. En linjär intonationskurva dras mellan värdena för grundtonsfrekvensen. Algoritmen MBROLA sammanfogar difonerna efter textfilens utseende och jämnar ut gränserna mellan varje segment. Slutligen genereras en ljudfil med 16 bitars syntetiskt tal (Dutoit m. fl.,1996).

Artikulatorisk talsyntes

Talgenerering genom att modellera talapparatens rörelser, sk. artikulatorisk syntes, har varit mindre framgångsrik. Modellen innehåller flera parametrar som beskriver artikulationen. För varje fonem finns ett målvärde lagrat för parametrarna. En korrekt modell av talapparaten är svår att fånga då den mesta informationen hämtats från tvådimensionella röntgenbilder. Modellen blir därför oftast förenklad och syntetiskt tal genererat från artikulatorisk syntes håller ofta en låg kvalitet (O'Shaughnessy, 2000:346f).

Formantsyntes

Mer framgångsrik är däremot sk. formantsyntes. Målet med formantsyntes är att producera de talljuden som ingår i mänskligt tal genom att skicka impulser från en ljudkälla genom filter och resonatorer. Syntesmodellens arkitektur är inte en modell av talproduktionen i talapparaten. För att människoliknande ljud ska kunna produceras används istället parametrar som sätts och justerar formanternas amplituder och bandbredd. Parametrarna förändrar signalen så att exempelvis ljud liknande nasaler, frikativor, klusiler och viskningar kan genereras. På grund av otillräckliga fonetiska kunskaper är framgångarna för formantsyntes begränsade (O'Shaughnessy, 2000:349ff).



Figur 2.1: Översikt över ett text-till-talsystem

2.2 Grunder i svensk prosodi

Prosodi är en samlingsbeteckning för de egenskaper i talet som uppfattas rytmiska, dynamiska och melodiska. Egenskaperna realiseras i talet genom ett samspel av de akustiska korrelerade duration, grundtonsfrekvens (F0) och intensitet (tryckstyrka). Variationerna av de tre akustiska korrelerade påverkar inte enbart varandra. Prosodins samtliga akustiska korrelerade påverkas såsom exempelvis talets ljudspektrum så att vokal- och konsonantkvaliteten förändras (Bruce, 1998:11).

2.2.1 Duration

Ett ljud har en inherent längd som är beroende av dess kvalitet. En öppen vokal har en längre duration än en sluten. Även omkringliggande ljud har en inverkan på ljudets duration. Artikulationssättet på en postvokal konsonant påverkar vokalens duration. Dessutom är en vokal framför en tonande konsonant längre än om den förekommer framför en tonlös. Ju längre durationen är för en pre- eller postvokal konsonant, desto kortare blir vokalen. Konsonanters inverkan på vokalers duration är dock inte alltför stor (Lindblom, Lyberg och Holmgren, 1976:19).

I svenskan råder dessutom komplementär distribution. I ordparet *vit* – *vitt* har vokalljudet *i* samt konsonantljudet *t* olika duration. En lång vokal följs av en kort konsonant och en kort

vokal följs av en lång konsonant (Lyberg, 1981:5). Durationen av en betonad vokal är en funktion av antalet föregående och efterföljande stavelser i ordet. Ju fler stavelser ett ord innehåller desto kortare duration får den betonade vokalen i ordet. Den betonade vokalen i ordet *dag* är durationsmässigt längre än den betonade vokalen *a* i ordet *Dagobert* (Lyberg, 1981:26).

2.2.2 Accenter och grundtonsfrekvens

Svenskan, liksom danskan och norskan, är ett sk. accentspråk. I svenskan finns det två typer av accenter som den huvudsakliga betoningen kan ha. Den ena är accent 1 (akut) och den andra accent 2 (grav). Den akustiska skillnaden mellan de två skildras främst genom tidsläget av grundtonsfrekvensens minimum och maximum, men även genom intensiteten och durationen (Lyberg, 1981:6).

Ordet *and+en* (fågel) bär accent 1 och har endast en prominent stavelse och en markant stigning av grundtonsfrekvensen i isolerat uttal medan *ande+n* (själ) bär accent 2 och har två prominenta stavelser och två markanta grundtonsfrekvenstoppar i isolerat uttal (Ladd, 2000:38). Uttalet är dialektalt och exemplet gäller företrädesvis då orden uttalats med stockholmsdialekt.

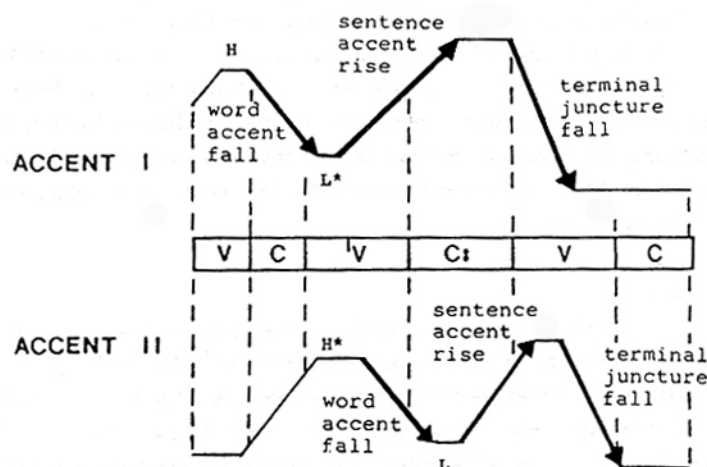
Accent 1

Grundtonsfrekvensens maximum för ett ord som bär accent 1 ligger i vokalen i stavelsen som föregår den accentuerade stavelsen, i den sk. pretoniska stavelsen. Från ordaccentens maximum sker en sänkning till grundtonsfrekvensens minimum. Grundtonsfrekvensens minimum ligger i början av vokalen i den accentuerade stavelsen. Om ingen pretonisk stavelse finns, exempelvis om den accentuerade stavelsen står satsinitialt eller fokalt, uteblir tontoppen i den pretoniska stavelsen och ordaccentsfallet. En tonal stigning sker med början i den betonade stavelsen till en tontopp senare i ordet, sk. satsaccent. Då ordet står fokalt och ordaccenten uteblir framträder endast satsaccenten. Det finala fallet i grundtonsfrekvens, sk. terminal junktur, i ett ord som bär accent 1 påbörjas normalt strax innan eller på konsonant-vokalgränsen i ordets sista stavelse, beroende på om konsonanten är betonad eller inte. Då ordet står fokalt terminerar fallet mitt i vokalen i ordets sista stavelse. I en icke fokal ställning förskjuts fallet och avslutas i slutet av vokalen eller i efterföljande konsonant (Bruce, 1998:104, 1977:20 och 37ff).

Accent 2

I ett ord som bär accent 2 ligger en tontopp i början av vokalen i den primärbetonade stavelsen och ett fall sker genom den betonade stavelsen, sk. ordaccent. Då ordet står i fokus eller isolerat tillkommer ytterligare en tontopp, sk. satsaccenten, mitt i vokalen i den sekundärbetonade stavelsen i ordet. Då det enskilda ordet inte står fokalt och i en kontext som i sin tur kan vara fokal, tonas satsaccenten i ordet ned och kan istället förskjutas och placeras på den (primär-) betonade stavelsen i meningens sista ord. Den terminala junkturen för ett ord som bär accent 2 påbörjas senare än för ett ord som bär accent 1. Fallet påbörjas mitt i vokalen i den andra betonade stavelsen och terminerar normalt i den senare delen av samma vokal. Liksom för ord som bär accent 1 förskjuts fallet något i en icke fokal ställning (Bruce, 1998:101 och 104, 1977:20, 58 och 37ff).

Den tonala stigningen, satsaccenten, förekommer i ord både med accent 1 och 2, men senare i det sistnämnda. Tontoppen för accent 2 är oftast högre än för accent 1. Däremot är ordaccentsfallet, då det finns, oftast brantare för accent 1 än accent 2 (Bruce, 1998:104).



Figur 2.2: Grundtonsförloppet för ordaccentuering, fokusering och frasfinalitet för ett accent 1 och ett accent 2 ord uttalat med stockholmsdialekt (Bruce, 1977:50).

Nivåtoner

Bruce (1977:132) formulerade en teoretisk notation för beskrivning av grundtonsfrekvensens relativa nivå över ord och yttranden. Nivåerna är H (hög) och L (låg) som kopplas samman genom linjär interpolering. H står för lokalt maximum av grundtonsfrekvensen medan L står för lokalt minimum. Beskrivningen har en stark koppling till den autosegmentella fonologin där en nivå med toner introducerats som skild från en segmentell och fonetisk nivå. Även inom den autosegmentella fonologin struktureras accent- och tonspråken med nivåtonerna H och L för att finna en länk mellan stavelser i ett yttrande och accenterna eller tonerna. För att indikera placeringen av det betonade segmentet (Frid, 1999) och markera att det finns en länk mellan ett stavelse- eller fonemsegment och en ton används en stjärna (*) som symbol (Bruce, 1998:44ff). En ton är däremot inte nödvändigtvis länkad till en specifik position i segmentnivån. En tons position är på så vis bestämd av dess relation till omkringliggande toner och inget specifikt segment i segmentnivån är kopplat till den tonen (Frid, 2003:78).

Bruces beskrivning av grundtonsfrekvensens nivåer ligger som grund till flera olika uttrycksätt för svenskans accent 1 och 2. Beskrivningen av grundtonsfrekvensens nivå är länkad till den accentuerade stavelsen. Fant och Kruckenberg använder notationen H L* Ha för att utmärka accentueringen i ett ord med accent 1 och H* L Hg för accent 2. Notationen för den huvudbetonade stavelsen i ett ord med accent 1 börjar med L*. H är placerad i den pretoniska stavelsen. Ha är placerad inom den accentuerade stavelsen och markerar en stigning av grundtonsfrekvensen, vilken är markant då ordet är fokalt. Den primärbetonade stavelsen i ett ord med accent 2 betecknas H*L. Då ordet är fokalt sker en markant sänkning av grundtonsfrekvensen i transitionen mellan H* och L. Den sekundärbetonade stavelsen i ett ord med accent 2 betecknas Hg och är placerad inom stavelsen (Fant, Kruckenberg, Liljencrants, 2000). De accentuerade stavelserna i ord som bär accent 1 eller 2 betecknas med två punkter och alla andra stavelser med endast en. Notationen Lu betecknar obetonade stavelser. Exempelordet *postlåda* representeras enligt följande: post - lå - da
H*L - Hg - Lu

(Frid, 2003:87)

2.2.3 Betoning

I svenskan finns två huvudsakliga ordbetoningsmönster; enkelt och sammansatt. Det enkla betoningmönstret kännetecknas av endast en primär betoning medan det sammansatta betoningmönstret kännetecknas av både en primär och en sekundär. Betoningen sträcker sig över en stavelse i ett ord och är distinktiv, vilket framgår av ordparen *formel* – *formell* samt *modern* – *modern* (Bruce, 1998:30f). Det finns flera sätt att märka upp betonade och obetonade stavelser. Ett sätt som i viss mån följer Svenska Akademiens Ordboks uppmärkning är att den accentuerade stavelsen i ett ord med accent 1 har beteckningen 4. I ett ord med accent 2 har den primärbetonade stavelsen beteckningen 3. Ett sammansatt ord har i regel accent 2 och den sekundärbetonade stavelsen i ett sammansatt ord har beteckningen 2. Ett ord som har accent 2 och inte är sammansatt har beteckningen 1, vilken i SAOB används i en vidare mening. En obetonad stavelse har beteckningen 0 (Fant, 2001).

2.3. Prominens och gruppering

Prosodin är suprasegmentell och påverkar en stavelse som minsta enhet ända upp till en talparagraf som största enhet. En talparagraf är en grupp yttranden som kan identifieras som prosodiskt samhörig. I talet fyller prosodin flera språkliga och kommunikativa funktioner. De två främsta är prominens (viktning) och gruppering. Prominens framhäver eller undanhåller en del av ett yttrande och gruppering markerar gränser. Genom ett samspel av båda funktionerna struktureras talet, vilket underlättar talkommunikationsprocessen (Bruce, 1998:11f, 15).

Prominens förnimmas genom att sticka ut från sin omgivning och är inte absolut utan beroende av sin miljö. Ett kontrastivt fokus påverkar alla ingående stavelser i yttrandet som i sin tur påverkar yttrandets prosodiska struktur och semantik. Ett fåtal modeller har utvecklats och ett värde för prominens har introducerats för att fånga det kontrastiva fokuset utifrån lexikala, syntaktiska och semantiska regler. I Fants och Kruckenbergers modell (1989) användes en skala mellan 0 och 31 där det mest prominenta ordet i ett yttrande tilldelades prominensnivån 31. Price et. al (1991) använde sig av en 7-gradig skala och Rietveld och Gussenhoven (1992/-93) av en skala från 1 till 100 för att mäta samma kvantitativa parameter (Portele & Heuft, 1997).

Den perceptuellt uppfattade prominensnivån relaterar i de flesta modellerna till en eller flera prosodiska parametrar; duration, paus, grundtonsfrekvens och i viss mån subglottalt tryck. Durationen av en betonad stavelse i ett ord ökar likväl som durationen av ett ords samtliga stavelser ökar då hela ordet är betonat. Samtidigt som durationen förlängs för ett prominent ord kan grundtonsfrekvensen stiga (Fant & Kruckenberg, 1999). Enligt Heldner (1998) är däremot en höjning av grundtonsfrekvensen för ett prominent segment varken nödvändigt eller tillräckligt som enda parameter för att segmentet eller uttrycket ska uppfattas som prominent. Ett enskilt ord eller yttrande föregås ofta av en kort paus kombinerat med ett terminalt fall av grundtonsfrekvensen för att framhäva dess prominens. Då ett uttryck uttalas med en mycket kraftig emfas ökar dessutom det subglottala trycket (Fant, Kruckenberg, Liljencrants, 2000).

Gruppering signalerar koherens och gränser i talet där samrådighet mellan språkets grammatiska enheter och de fonetiska och fonologiska inte alltid går att finna. Överensstämmelsen mellan syntaktisk och prosodisk frasstruktur är bättre högre upp i den syntaktiska hierarkin. Syntaktiska grupper som satser inom en mening signaleras ofta även prosodiskt till skillnad mot djupare inbäddade syntaktiska konstituenten som exempelvis en prepositionsfras inbäddad i en nominalfras (Bruce, 1998:124 och 127ff).

Temporala relationer mellan prosodiska segment är betydelsefulla för att signalera gränser i talet. En paus i talet stoppas gärna in mellan grupper högt upp i den syntaktiska hierarkin, exempelvis vid meningsslut och paragrafslut. Längre durationer av finala fonem och stavelser, sk. final förlängning samt att slå av på talhastigheten är också gränssignalerande. En sänkning av grundtonsfrekvensen i slutet av en prosodisk grupp signalerar ett avslutande samtidigt som grundtonsfrekvensen och intonationen återställs till ett högre läge i början av en ny prosodisk grupp. Sänkning i intensitet och röstkvalitet, exempelvis knarr, kan också markera en prosodisk gräns. Koherens signaleras i talet genom frånvaron av de fonetiska korrelaten som är gränssignalerande (Bruce, 1998:132f).

2.4 FK-systemet

Prosodimodellen FK-systemet består av prosodiregler för svenska med stockholmsdialekt och avses att inkorporeras i ett text-till-talsystem. Materialet för utvecklingen av modellen är en studie av tal från fem testpersoner. Två av testpersonerna var kvinnor och tre var män. Testpersonerna läste avsnitt ur en svensk roman under tre minuter. Reglerna konstruerades utifrån studiernas resultat. För att få kvantitativt likvärdiga intonationskurvor för kvinnorna och männen tidsnormaliserades intonationsförloppet samt frekvensskalan normaliserades så att grundtonsfrekvensen representerades på en halvtonsskala (Fant m. fl., 2002). Tillägg till reglerna har skett efter generering av ett flertal exempelmeningar från nyhetstexter.

För att reglerna ska kunna tillämpas på texten kräver modellen vissa förutsättningar. Ur texten måste paragrafer, meningar, huvudsatser, bisatser, nominalfraser, prepositionsfraser och ord kunna härledas. Varje ord har en ordklassstillhörighet och det måste framgå om ordet har accent 1 eller 2. Ur orden måste stavelser kunna härledas och därur fonem samt huruvida stavelserna är obetonade, primär- eller sekundärbetonade i ett sammansatt ord eller i ett icke sammansatt ord (Fant & Kruckenberg, 2001).

2.4.1 Prominens

I FK-systemet ingår en kontinuerligt skalad prominensparameter, även kallad R_s , som påverkar både grundtonsfrekvensen och fonemens durationsvärden. Ett ords prominensvärde bestäms av dess ordklassstillhörighet. Ju högre värdet är desto mer prominent är ordet. Värdet kan komma att modifieras beroende på om ordet är fokalt. Ett innehållsord har ett prominensvärde över 14 medan funktionsord har 14 och lägre. Ett fokalt ord har ett prominensvärde högre än 22,5 (Fant & Kruckenberg, 2001). Till innehållsorden räknas ord som har en egen semantisk betydelse exempelvis substantiv, adjektiv och verb (ej hjälpverb). Däremot kan även stavelser i funktionsord betonas då ordet står fokalt och får därmed en högre prominens (Werners & Keller:1994). I FK-systemet tilldelas interjektion prominensvärdet 24, adjektiv 21,5, räkneord 21, substantiv 20,5, verb 18,5, determinativt pronomen 17, adverb 15, hjälpverb 12, pronomen 12 och resterande ordklasser 11 (Fant & Kruckenberg, 2001).

Förutom att prominensparametern R_s sätts för varje ord sätts även en prominensparameter, även kallad R_{sF0} , för varje beteckning av grundtonsfrekvensens nivå; H, L*, H*, L, Hg, Ha och Lu. Parametern sätts till samma värde som ordets prominensvärde för alla nivåerna med undantag för de som betecknats som obetonade, dvs. Lu, och även för H. För Lu sätts parametern R_{sF0} till 11 och för H sätts den till det ordet accenten ingår i, dvs till samma värde som för L* och Ha. R_{sF0} används i beräkningen av den faktiska grundtonsfrekvensen. En mindre prominent stavelse ges ett lägre prominensvärde (Fant, 2002).

Ytterligare en prominensparameter finns i systemet, R_{sD} , och sätts för varje fonem. Alla

fonem inom samma stavelse tilldelas samma värde på parametern RsD. Värdet sätts till samma som ordets värde på parametern Rs förutom i några fall där föregående eller efterföljande betonade stavelser påverkar värdet av RsD för efterföljande eller föregående fonem (Fant & Kruckenberg, 2001).

2.4.2 Gruppering

De prosodiska grupperna skiljs åt med junkturer som paus, final förlängning och en initial stigning av grundtonsfrekvensen. I FK-systemet tilldelas större prosodiska grupper som huvudsatser, bisatser, längre nominal- och prepositionsfraser en intonationsmodul på vilken accent 1 och accent 2 modulering överlagras. En sådan modul kallas baskurva. Systemet består av fyra olika typer av baskurvor varav två kan komma att modifieras beroende på om baskurvan är sist i en paragraf eller text. Typen ob1a samt ob1b inleder en ny mening. De baskurvorna har en initial stigning i grundtonsfrekvens som därefter sjunker till $-1,5$ halvtoner. Ob1b avslutar även en mening. Typen ob2a och ob2b tar vid inne i en mening. Kurvan saknar en markant initial stigning, men har en markant sänkning av grundtonsfrekvensen och intonationsmodulens sista stavelse får en grundtonsfrekvens på -4 halvtoner. Om intonationsmodulen är den sista innan ett paragraf- eller textslut sänks grundtonsfrekvensen med 2 halvtoner för hela intonationsmodulen. Beroende på om en mening består av endast en kurva eller flera är typen av baskurva ob3a eller ob3b (Fant & Kruckenberg, 2001).

2.4.3 Junkturer och pauser

En junktur är en gräns mellan prosodiska enheter oavsett om den innehåller en paus eller inte. I regel består junkturen av en kombination av paus, final förlängning, förändring av intensitet och grundtonsfrekvens. I språk där regelbundna pauser observerats finns två sorters pauser, tysta och fyllda pauser. Med begreppet tyst paus avses ett avbrott i talsignalen som inte utgör en del av ljudstrukturen. En fylld paus markeras enbart med final förlängning och lokala förändringar i röstbildningen (Barbosa Ferreira, 2003).

Talsegmentens längd påverkar pausens längd. En lång mening föregår en längre paus än en kort mening. Viktigt att påpeka är att pausens längd även är talarberoende (Fant, Kruckenberg & Barbosa Ferreira, 2003). Pausernas förekommande och längd är dessutom till viss del beroende av hur ofta de förekommer i en sats samt satsens syntaktiska konstituenten. Pauserna ska inte förekomma för tätt för att syntetiskt tal skall låta naturligt. FK-systemet genererar tysta pauser inom en mening beroende på avståndet till tidigare pauser. Pauserna som genereras är av skild duration och förekommer vid huvudsatsslut, före kommativering, bisatsslut som följs av huvudsats eller bisats och före eller efter prepositions- och nominal fras beroende på var i satsen frasen står. En tyst paus genereras alltid vid meningsslut och paragrafslut. Durationen är beroende av antalet stavelser i satsen, frasen eller meningen (Fant, 2003b).

2.4.4 Final förlängning

Ord i frasfinal ställning har en längre duration än ord i initial och medial ställning. En förlängning av stavelser i ett ord kan medföra en fokal signalering såväl som junktur. Förlängningen av ett frasfinalt fokalt ord beror både på ordets emfas och på junktursignaleringen (Heldner och Strangert, 2001). I FK-systemet signaleras varje junktur med final förlängning. Inte enbart de prosodiska grupperna som markerats med en egen baskurva föregås av final förlängning, utan även kortare nominal- och prepositionsfraser. Förlängningen sker på den sista stavelsens samtliga fonem. Då stavelsen är betonad samt meningssfinal sker en förlängning med 30% på stavelsens samtliga fonem. I övriga fall där

final förlängning sker förlängs fonemen i den sista stavelsen med 60%. Förlängningen sker vare sig stavelsen är betonad eller obetonad (Fant, 2003a).

2.4.5 Accentmodulering

Ett grundtonsfrekvensvärde räknas ut för varje nivå i accentmoduleringen. Moduleringen av ordaccent, vilken bygger på Bruces (1977) beskrivning, placerar nivåerna på bestämda platser inom en accentdomän, dvs. det ordet eller de orden som påverkas av ett ords accent.

Notationen L^* placeras på vokalens vänstra gräns i den betonade stavelsen i ett accent 1 ord. H placeras på samma vokals högergräns, eller efterföljande konsonants högergräns om den efterföljande konsonanten är tonande och inte klusil. H placeras mitt i vokalen i närmast föregående obetonade stavelse inom samma prosodiska grupp som markeras med byte av baskurva, om där finns någon. Om en obetonad stavelse inte finns utblir H .

H^* placeras på vokalens vänstergräns i den primärbetonade stavelsen i ett accent 2 ord. L placeras 150 millisekunder efter H^* , men inte senare än på samma fonem som Hg är placerad på. Hg placeras i den sekundärbetonade stavelsen i ett accent 2 ord på vokalens högergräns eller på efterföljande konsonants högergräns om den efterföljande konsonanten är tonande och inte klusil. I modellen delas Hg upp i $Hg1$ och $Hg2$. $Hg1$ betecknar den sekundärbetonade stavelsen i ett icke sammansatt ord och $Hg2$ i ett sammansatt.

Lu placeras mitt i vokalen i obetonade stavelser. Ord med ett prominensvärde lägre än 15 klassas varken som accent 1 eller 2 ord. Varje stavelse räknas då som obetonad och får notationen Lu (Fant & Kruckenberg, 2001). Se 3.3 *Accentmodulering* för vidare förklaring.

2.4.6 Beräkning av grundtonsfrekvens

Varje nivå i accentmoduleringen tillskrivs ett grundtonsfrekvensvärde. Flera faktorer är avgörande för utseendet av grundtonsfrekvensens kurva över ett yttrande. Varje grundtonsfrekvensvärde beräknas utifrån den relativa positionen i baskurvan, typ av baskruva, nivåbeteckning och värdet på prominensparametern R_sF_0 . Formlerna som tillämpas vid beräkningen av grundtonsfrekvensen skiljer sig beroende på vilken nivåbeteckning placeringen för grundtonsfrekvensen har samt i vilken typ av baskurva den finns (Fant & Kruckenberg, 2001).

2.4.7 Fonemduration

För beräkning av fonemdurationerna används en statistisk databas bestående av fonems duration i millisekunder i en betonad respektive obetonad stavelse. Hänsyn har även tagits till kontexten i stavelsen. Se *Appendix* för fonemdurationer. Ett fonems exakta duration påverkas av värdet på prominensparametern R_sD hos fonemet, vilken kontext fonemet är placerat i samt fonemets duration både i en obetonad respektive betonad stavelse. Förutom final förlängning som påverkar den beräknade durationen av ett fonem hanterar systemet även beräkningar för reduktion vid konsonantföljder. Durationen för varje konsonantljud då flera konsonanter följer efter varandra påverkas. Durationen påverkas då konsonantljuden förekommer inom en och samma stavelse, över stavelsegränserna och även över ordgränserna. Däremot påverkas den inte över gränsen för en ny intonationsmodul (Fant & Kruckenberg, 2001).

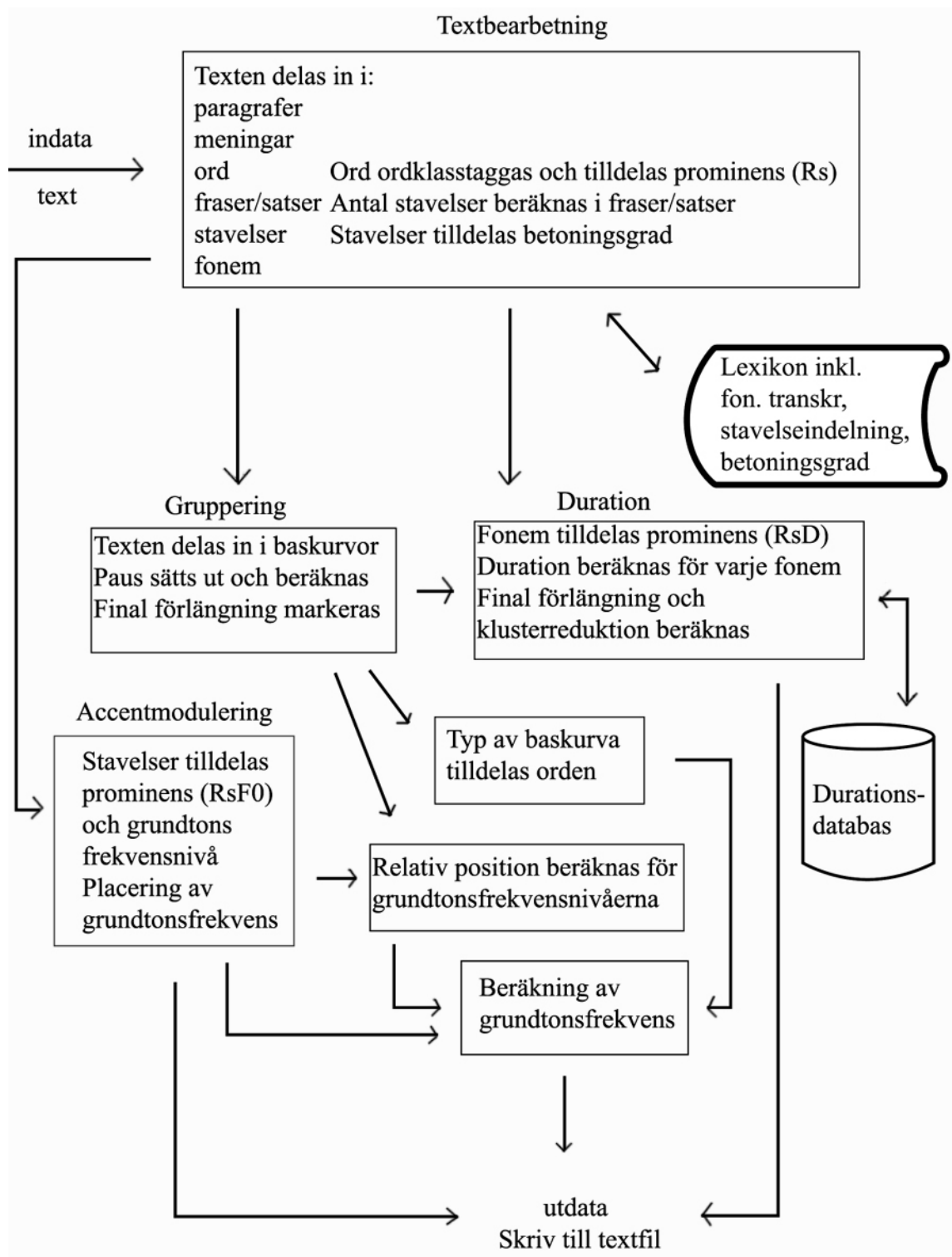


Bild 2.3: Illustration av FK-systemets ingående delar

3 Genomförande

Utifrån FK-systemets ursprungliga skick i form av dokumentation, grundtons- och durationsberäkningar av exempelmeningar i kalkylprogrammet Excel och ett flertal diskussioner med professor Gunnar Fant har en struktur och förslag till implementering av FK-systemet arbetats fram. Implementeringen är objektorienterad och skriven i pseudokod för att enkelt kunna omskrivas till ett objektorienterat programmeringsspråk. Inledningsvis studerades uppbyggnaden av text-till-talsystemet Festival utvecklat av The Centre for Speech Technology Research (CSTR) vid Edinburghs universitet i Skottland för att undersöka möjligheterna till en implementering i Festival. Då FK-systemet är ett system som utvecklas kontinuerligt lämnades integreringen av FK-systemet i Festival därhän och fokus riktades på en utarbetning av FK-systemets mindre definierade delar för att slutligen utveckla en tillämpbar algoritm.

Nedan följer en beskrivning av implementeringen av FK-systemets mest centrala delar: Textbearbetning, gruppering, accentmodulering, beräkning av grundtonsfrekvens och beräkning av duration.

3.1 Textbearbetning

FK-systemet är modulärt. I systemets första fas definieras textens beståndsdelar utifrån de förutsättningar modellen kräver. Modellen kräver att texten delas in i meningar, fraser, ord, stavelser och fonem samt att varje grundtonsfrekvensnivå markeras. Varje beståndsdel utgör en egen datastruktur. Datastrukturerna är hierarkiskt uppdelade och därmed nåbara från en struktur på en högre nivå.

Varje datastruktur består av en sträng tecken, exempelvis hela texten, en mening, en fras, ett ord, en stavelse eller ett fonem och är kallad därefter. Var och en av datastrukturerna, förutom den som definierar fraser och den som definierar grundtonsfrekvensnivån, består av en lista av en lägre stående datastruktur, dvs. datastrukturen *Mening* består av en lista där varje element i listan är av datastrukturen *Ord* och listan består av meningens alla ord, ord består av stavelser, stavelser av fonem och fonem av inga till maximalt två grundtonsfrekvensnivåer. Varje datastruktur har dessutom strukturspecifika attribut.

Datastrukturerna *Mening* och *Ord* består även av en variabel som lagrar antalet stavelser i meningen respektive ordet. *Mening* består dessutom av en variabel för huruvida meningen är sist i en paragraf. I datastrukturen *Ord* lagras värden för om ett skifte av baskurva sker efter ordet, typ av baskurva ordet tillhör, om tilldelning av paus sker efter ordet, ordets prominensvärde (R_s), ordklass och om den sista stavelsen ska ha final förlängning. Förutom att *Ord* dessutom består av stavelser består den av datastrukturen *Fras* som definierar vilka fraser ett ord ingår i. I datastrukturen *Stavelse* lagras stavelsens betoningsgrad 0-4. I datastrukturen *Fonem* lagras fonemets duration i en variabel samt fonemets prominensvärde (R_sD).

Förutom en sträng tecken består datastrukturen *Fras* av frasens typ som det ordet vi befinner oss på ingår i, exempelvis prepositionsfras, nominalfras etc. FK-systemets funktion för indelning av fraser kräver att prepositionsfras, nominalfras, huvudsats, bisats, ”enkel uppräknings” och ”komplex uppräknings” kan identifieras. En sk. ”enkel uppräknings” består av en uppräknings av nominal- eller adjektivfraser bestående av endast ett ord med kommatecken och/eller konjunktion emellan, exempelvis Lisa, Pelle och Lena. En sk. ”komplex uppräknings” består av en uppräknings av nominal- eller adjektivfraser bestående av fler än ett

ord med kommatecken och/eller konjunktion emellan, exempelvis "Lisas hund, Pelles katt och Lenas häst" där varje ingående nominalfras, ex "Lisas hund" ska kunna identifieras. I datastrukturen *Fras* lagras även antalet stavelser i frasen samt index i meningen för frasens första ord och index i meningen för frasens sista ord. Exempel: "Hans hund sprang".
 Fraslistan för "Hans": [{0,0,"np"},{0,1,"np"}, {0,2,"huvudsats"}]. Fraslistan för "hund": [{0,0,"np"},{0,1,"np"},{0,2,"huvudsats"}]. Fraslistan för sprang: [{0,2,"huvudsats"}]. FK-systemet kräver inte att samtliga fraser i en text ska kunna identifieras. Det är exempelvis inte nödvändigt att kunna identifiera en verbfras (Fant, 2003b).

Datastrukturen *Punkt* består av en variabel som lagrar grundtonsnivåns prominensvärde (RsF0), grundtonsfrekvens, semiton, index i baskurvan och typ, dvs. beteckning på grundtonsfrekvensnivån: H, L*, H*, L, Hg1, Hg2, Ha och Lu. Dessutom lagras ett värde i procent som anger var i fonemet nivån uppnås där 0 procent är längst till vänster på fonemet, dvs. vid duration 0 av fonemet. Då placeringen av en grundtonsfrekvensnivå är beroende av den reella placeringen på den föregående nivån lagras värdet i millisekunder i variabeln avstånd. Punkten L förekommer 150 millisekunder efter punkten H* och är därmed inte styrd av en förutbestämd placering på fonemet angiven i procent.

3.2 Gruppering

Gränssignalerande i FK-systemet är insättning av paus, final förlängning på ett ords sista stavelse samt skifte av baskurva. Tillämpningen av de tre momenten är starkt sammanknutna. Tilldelningen av paus, final förlängning och skifte av baskurva är beroende av stavelseantal i en sats eller fras och satsens eller frasens placering i meningen. Då satser eller fraser står i varandra, exempelvis "huset på berget", ska den längsta och mest omfattande satsen eller frasen studeras. I exemplet "huset på berget" beaktas hela nominalfrasen och inte prepositionsfrasen "på berget" som är en del av den längre nominalfrasen. Tilldelningen är indelad i sex pass för att hålla isär de villkor som måste uppfyllas för att tilldelningen ska kunna ske. Alla sex passen utförs i nummerordning. Då regler för insättning av paus, final förlängning och markering av skifte av baskurva sammanfaller skrivs tidigare uppmärkning över med en senare utförd regel. Slutligen sker en tilldelning av enbart final förlängning som gränssignalerande.

Pass 1 hanterar villkoren då kommatecken uppstår i texten:

Om den aktuella meningen har en huvudsats följt av kommatecken och efterföljande ord är början av en huvudsats eller om den aktuella meningen har ett kommatecken och det är minst 8 stavelser till föregående paus/skifte (exklusive enkel uppräknings) sker:
 skifte av baskurva efter ordet som föregår kommatecknet
 final förlängning på ordet som föregår kommatecknet
 sätt paus till 450 ms efter ordet som föregår kommatecknet
 Om det aktuella ordet inte är det sista i en huvudsats:
 sätt paus till 175 ms efter ordet som föregår kommatecknet

Pass 2 hanterar villkoren då huvudsatser och bisatser följer varandra:

Om den aktuella meningen har en huvudsats följt av en huvudsats och en konjunktion kan förekomma mellan huvudsatserna sker:
 Skifte av baskurva efter det sista ordet i den föregående huvudsatsen
 final förlängning på det sista ordet i den föregående huvudsatsen
 Om den föregående huvudsatsen består av minst 13 stavelser:
 sätt paus till 450 ms efter det sista ordet i den föregående huvudsatsen
 Annars om den föregående huvudsatsen består av minst 8 och max 12 stavelser:
 sätt paus till 175 ms efter sista ordet i den föregående

huvudsatsen
Annars om den föregående huvudsatsen består av minst 2 och max 7 stavelser:
sätt paus till 75 ms efter det sista ordet i den föregående huvudsatsen

Om den aktuella meningen har en huvudsats följt av en bisats, bisats följt av en huvudsats eller bisats följt av en bisats och en konjunktion kan förekomma mellan satserna, satserna tillsammans utgör minst 20 stavelser, det är minst 8 stavelser till föregående skifte av baskurva, den föregående satsen är minst 8 stavelser och det är minst 8 stavelser till efterföljande skifte av baskurva sker:

skifte av baskurva efter det sista ordet i den föregående huvud- eller bisatsen
final förlängning på det sista ordet i den föregående huvud- eller bisatsen

Om den föregående huvud- eller bisatsen består av minst 19 stavelser:
sätt paus till 450 ms efter det sista ordet i den föregående huvud- eller bisatsen

Om den föregående huvud- eller bisatsen består av minst 13 och max 18 stavelser:

sätt paus till 175 ms efter det sista ordet i den föregående huvud- eller bisatsen

Om den föregående huvud- eller bisatsen består av minst 8 och max 12 stavelser:

sätt paus till 75 ms efter det sista ordet i den föregående huvud- eller bisatsen

Pass 3 hanterar villkoren då längre prepositions- eller nominalfraser förekommer i en mening:

Om den aktuella meningen har en prepositions- eller nominalfras som står i en huvud- eller bisats som består av minst 20 stavelser, det är minst 8 stavelser till efterföljande skifte av baskurva och prepositions- eller nominalfrasen består av minst 8 stavelser sker:

Om prepositions- eller nominalfrasen påbörjas färre än 8 stavelser från början på huvud- eller bisatsen sker:

skifte av baskurva efter det sista ordet i prepositions- eller nominalfrasen

final förlängning efter det sista ordet i prepositions- eller nominalfrasen

Om prepositions- eller nominalfrasen består av minst 13 stavelser:
sätt paus till 175 ms efter det sista ordet i prepositions- eller nominalfrasen

Om prepositions- eller nominalfrasen består av minst 8 och max 12 stavelser:

sätt paus till 75 ms efter det sista ordet i prepositions- eller nominalfrasen

Annars om prepositions- eller nominalfrasen påbörjas minst 8 stavelser från början på huvud- eller bisatsen sker:

Om ordet som föregår prepositions- eller nominalfrasen är en konjunktion och konjunktionen tillsammans med prepositions- eller nominalfrasen utgör minst 8 stavelser och det är minst 8 stavelser till föregående skifte av baskurva sker:

skifte av baskurva efter ordet som föregår konjunktionen

final förlängning på ordet som föregår konjunktionen

Om prepositions- eller nominalfrasen inklusive konjunktionen tillsammans utgör minst 13 stavelser:

Sätt paus till 175 ms efter ordet som föregår konjunktionen

Om prepositions- eller nominalfrasen inklusive konjunktionen tillsammans utgör minst 8 och max 12 stavelser:

Sätt paus till 75 ms efter ordet som föregår konjunktionen

Annars om ordet som föregår prepositions- eller nominalfrasen inte är ett verb och det är minst 8 stavelser till föregående skifte av och prepositions- eller nominalfrasen består av minst 8 stavelser sker:

skifte av baskurva efter ordet som föregår prepositions- eller nominalfrasen

final förlängning på ordet som föregår prepositions- eller nominalfrasen

Om prepositions- eller nominalfrasen består av minst 13 stavelser:

Sätt paus till 175 ms efter ordet som föregår prepositions- eller nominalfrasen
Om prepositions- eller nominalfrasen består av minst 8 och max 12 stavelser:
Sätt paus till 75 ms efter ordet som föregår prepositions- eller nominalfrasen

Pass 4 hanterar villkoren då ”komplex uppräknig” och ”enkel uppräknig” förekommer i en mening.

Om den aktuella meningen har en komplex uppräknig sker:
Skifte av baskurva efter det sista ordet i den första ingående nominalfrasen
final förlängning på det sista ordet i den första ingående nominalfrasen
paustilldelning med 175 ms efter det sista ordet i den första ingående nominalfrasen

Om den aktuella meningen har en komplex uppräknig sker:
final förlängning på det sista ordet i den sista ingående nominalfrasen
Om den komplexa uppräknigens består av minst 13 stavelser:
sätt paus till 175 efter det sista ordet i den komplexa uppräknigens
Om den komplexa uppräknigens antal stavelser består av minst 8 och max 12:
sätt paus till 75 efter det sista ordet i den komplexa uppräknigens

Om den aktuella meningen har en komplex uppräknig sker:
skifte av baskurva efter det sista ordet i de resterande ingående nominalfraserna
final förlängning på det sista ordet i de resterande ingående nominalfraserna
paustilldelning med 450 ms efter det sista ordet i de resterande ingående nominalfraserna

Om den aktuella meningen har en enkel uppräknig sker:
final förlängning på varje ord
paustilldelning med 75 ms efter varje ord
Om den enkla uppräknigens består av minst 13 stavelser:
sätt paus till 175 ms efter det sista ordet i den enkla uppräknigens
Om den enkla uppräknigens består av minst 8 och max 12 stavelser:
sätt paus till 75 m efter det sista ordet i den enkla uppräknigens

Pass 5 hanterar villkoren då en mening avslutas:

Sätt paus till 10 * antalet stavelser i meningen + 850 ms efter det sista ordet i en mening
Sätt final förlängning på meningens sista ord

Pass 6 hanterar villkoren då en paragraf avslutas.

Sätt paus till 1500 ms efter det sista ordet i en paragraf

Vid följande fall förkommer enbart final förlängning för att gruppera uttrycket.

Om den aktuella meningen har en huvud- eller bisats sker:
final förlängning på det sista ordet i huvud- eller bisatsen

Om den aktuella meningen har en prepositions-, nominalfras eller komplex uppräknig som består av minst 4 stavelser sker:
final förlängning på det sista ordet i prepositions-, nominalfrasen eller den komplexa uppräknigens

Om den aktuella meningen har en huvud- eller bisats och en huvudsats- eller bisatsinitial prepositions-, nominalfras eller komplex uppräknig sker:
final förlängning på det sista ordet i prepositions-, nominalfrasen eller den komplexa uppräknigens

3.3 Accentmodulering

Accentmoduleringen innebär att svenskans accent 1 och 2 samt obetoning markeras och placeras ut på stavelserna i en text. Accentmoduleringen sker i två funktioner där den första hanterar grundtonsfrekvensnivåerna Lu, H, L*, Ha, H*, Hg1 och Hg2. Varje grundtonsfrekvensnivå bär på ett grundtonsfrekvensvärde som ska uppnås. Maximalt två punkter per fonem kan förekomma. Varje punkt har en typ (Lu, H, L*, Ha, H*, Hg1, Hg2 eller L) och en procentsats som avgör hur långt in i fonemet grundtonsfrekvensen ska uppnås där 0 är längst till vänster på fonemet. Parametern RsF0 (prominens_F0) sätts och används senare för att beräkna grundtonsfrekvensen.

```
accentuering(Mening m) {
  för varje ord m.orden[i] gör {
    if m.orden[i].ordattribut.prominens < 15 {
      för varje stavelse m.orden[i].stavelser[j] gör {
        för varje fonem m.orden[i].stavelser[j].fonem[k] gör {
          if vokal(m.orden[i].stavelser[j].fonem[k]) {
            Punkt p
            p.typ=Lu
            p.procent=50
            p.avstånd=0
            p.prominens_F0=11
            m.orden[i].stavelser[j].fonem[k].punkter+=p
          }
        }
      }
    }
  }
}
else {
  för varje stavelse m.orden[i].stavelser[j] gör {
    if (m.orden[i].stavelser[j].stavelseattribut.betoningsgrad=3) {
      för varje fonem m.orden[i].stavelser[j].fonem[k] gör {
        if (vokal(m.orden[i].stavelser[j].fonem[k])) {
          Punkt p
          p.typ=H_star
          p.procent=0
          p.avstånd=0
          p.prominens_F0=m.orden[i].ordattribut.prominens_ord
          m.orden[i].stavelser[j].fonem[k].punkter+=p
          p.typ=L
          p.procent=0
          p.avstånd=150
          p.prominens_F0=m.orden[i].ordattribut.prominens_ord
          m.orden[i].stavelser[j].fonem[k].punkter+=p
        }
      }
    }
  }
}
```

Då typen är L anger attributet avstånd vart grundtonsfrekvensen ska uppnås och inte attributet procent. Attributet avstånd är avståndet i millisekunder och är 0 i de fall då procent istället används för att beräkna avståndet. I de fall då typen är L är avståndet 150 millisekunder. Punkten ska placeras 150 millisekunder efter placeringen av H*, men inte på ett senare fonem än där grundtonsfrekvensen Hg1 eller Hg2 är placerad. Funktionen för placering av typen L anropas efter att durationen beräknats för varje fonem i texten och innan grundtonsfrekvensen räknats ut. Funktionen räknar ut på vilket fonem grundtonsfrekvensnivån L ska placeras.

```
placering_L(Ord ord) {
  för varje stavelse ord.stavelser[i] gör {
    för varje fonem ord.stavelser[i].fonem[j] gör {
      if (ord.stavelser[i].fonem[j].punkter[1].typ = L &
          ord.stavelser[i].fonem[j].punkter[1].procent = 0) {
        Int stavelseindex = i
        Int fonemindex = j
        Int summa = 0
      }
    }
  }
}
```


RsD för ett fonem med grundtonsfrekvensnivån Hg2 beräknas:
 $(1,5 * prominens_dur\ för\ H^*) - 11,3$

RsD för ett fonem med grundtonsfrekvensnivån Lu som efterföljer Hg1, Hg2 eller i samma ord beräknas:
 $prominens_dur\ för\ Ha,\ Hg1\ eller\ Hg2 - 5$

RsD för ett fonem med grundtonsfrekvensnivån H är 11.
RsD för ett fonem med grundtonsfrekvensnivån Lu är 11.

För varje fonem i varje stavelse:

```
Om Punkt.typ = Hg1
  Prominens_dur = 0,5 * H* + 6,1
  Fonemdurations = (durationsvärdet för fonemet i en
    obetonad stavelse) + (((prominens_dur - 11) / 9) *
    durationsvärdet för fonemet i en obetonad stavelse -
    durationsvärdet för fonemet i en betonad stavelse))
```

RsD för sammansatta ord förändras för ett lexikalt obetonat ord. Ett lexikalt obetonat ord är ett eget ord som ingår i en sammansättning men saknar betoning i sammansättningen. Ett lexikalt obetonat ord får en annan prominens för durationen (RsD) än vad resten av sammansättningen får. Det lexikalt obetonade ordet tilldelas värdet på RsD efter ordets egen prominens (Rs) - 3 enheter. En extern funktion identifierar sammansättningens ingående ord med hjälp av ett sammansättningslexikon och tilldelar värdet på RsD för ett sammansatt ord. Exempelvis ordet *lastbilsflak* delas in i orden *last – bil(s) – flak*. *Bil(s)* som ett lexikalt ord har prominens 20,5. I sammansättningen tilldelas ordet värdet $20,5 - 3 = 17,5$ medan *last* och *flak* är betonade stavelser och tilldelas RsD 20,5.

3.5 Beräkning av grundtonsfrekvens

Grundtonsfrekvensen beräknas för varje grundtonsfrekvensnivå utifrån den relativa positionen på baskurvan, RsF0 och vilken baskurva grundtonsfrekvensnivån befinner sig i. I ett fall (L*) är även frekvensen för efterföljande grundtonsfrekvensnivå (Ha) nödvändig för beräkningen. Den relativa positionen för en grundtonsfrekvensnivå är dess ordningstal i baskurvan dividerat med summan av antalet grundtonsfrekvensnivåer i baskurvan. Beräkningen av grundtonsfrekvensen sker först i semitoner och räknas sedan om till Hertz för att anpassas till talsyntesens indata.

För beräkningen av grundtonen för baskurvorna används en formel för varje typ av baskurva. Utöver de formlerna tillämpas formler för varje grundtonsfrekvensnivå, förutom Lu. Formlerna för grundtonsfrekvensnivåerna är desamma för baskurvorna av typen ob1a, ob1b och ob3a medan en annan formel används för beräkningen av grundtonen för baskurvorna av typen ob2a, ob2b och ob3b. Frekvensen för grundtonsfrekvensnivån Lu beräknas endast utifrån det beräknade värdet för baskurvan. Då en obetonad stavelse (Lu) närmast efterföljer en stavelse i samma ord där grundtonsnivån är L, Ha, Hg1 eller Hg2 ärver däremot den obetonade stavelsen den föregående grundtonsnivåns hela värde för grundtonsfrekvensen. Se 4.6 *Beräkning av grundtonsfrekvens* för beräkning och exempel på formler.

Om baskurvan är Ob1a, Ob1b eller Ob3a:

För varje punkt i varje fonem i varje stavelse:

Om punkten ligger i en Ob1a:

Beräkna värdet för punkten enligt formeln för ob1a, Ob1b och ob3a

Om baskurvan är av typen ob1a:

Beräkna värdet för baskurvan enligt formeln för ob1a

Addera beräkningen av värdet för punkten till beräkningen av

värdet för baskurvan
 Om punkten ligger i en Ob1b:
 Beräkna värdet för punkten enligt formeln för ob1a, Ob1b och ob3a
 Om baskurvan är av typen ob1b:
 Beräkna värdet för baskurvan enligt formeln för ob1b
 Addera beräkningen av värdet för punkten till beräkningen av värdet för baskurvan
 Om punkten ligger i en Ob3a:
 Beräkna värdet för punkten enligt formeln för ob1a, Ob1b och ob3a
 Om baskurvan är av typen ob3a:
 Beräkna värdet för baskurvan enligt formeln för ob3a
 Addera beräkningen av värdet för punkten till beräkningen av värdet för baskurvan
 Subtrahera 2 från det adderade värdet

Då varje grundtonsfrekvensvärde tillskrivits en grundtonsfrekvensnivå räknas värdena i semitoner om till Hertz för att anpassas till talsyntesens indata. Formeln för beräkningen är:

$$Hertz = ((1.0595^{Punkt.semiton}) * 100)$$

4 Resultat

Som ett resultat av implementeringen av FK-systemet har ett exempel gått igenom. Exemplets utdata har genererats förhand utifrån definierade regler och förutsättningar. Meningen *Efter nästan sex år i Grekland for Anjalis till Rom, inbjuden av Petronius att föreläsa för stoikerna vid forum.* är tagen ur boken "Den som vandrar om natten..." av Marianne Fredriksson (1988:149).

4.1 Textbearbetning

Efter att texten delats in i meningar, ord, fraser och stavelser tilldelas orden ordklass samt prominensvärde (Rs). Stavelser tilldelas en betoningsgrad (0-4) med hjälp av lexikon.

Efter	nästan	sex	år	i	Grekland	for	Anjalis	till	Rom,
adv.	adv.	räkn.ord	subst.	prep.	subst.	verb	subst.	prep.	subst.
15	15	21	20,5	11	20,5	18,5	20,5	11	20,5

inbjuden	av	Petronius	att	föreläsa	för	stoikerna	vid	Forum.
adj.	prep.	subst.	inf.märke	verb	prep.	subst.	prep.	subst.
21,5	11	20,5	11	18,5	11	20,5	11	20,5

Ef – ter näs – tan sex år i Grek – land for An – ja – lis till Rom,
3 1 3 1 4 4 0 4 0 4 4 0 0 0 4

in – bju – den av Pe – tro – ni – us att fö – re - lä – sa för sto – i – ker – na vid Fo – rum.
3 1 0 0 0 4 0 0 0 3 0 1 0 0 4 0 0 0 0 3 1

4.2 Tilldelning av paus och final förlängning samt skifte av baskurva

I meningen förekommer ett skifte av baskurva. Det förekommer efter *Rom* efter att ha uppfyllt kriterierna för pass 1. Kommatecknet samt att det är minst 8 stavelser till föregående skifte föranleder skifte av baskurva. Vid skiftet förekommer även en paus och final förlängning av samtliga fonem i ordet *Rom*. Eftersom kommatecknet både föregås och efterföljs av huvudsatser sätts pausen till 450 millisekunder.

Efter hela meningen sker ett skifte av baskurva. En paus på 1220 millisekunder sätts ut efter meningen efter att ha uppfyllt kriterierna för pass 5. Meningen består av totalt 36 stavelser. $10 * 36 + 850$ ger 1220 millisekunders paus. Meningens sista ord, *forum*, tilldelas final förlängning.

Final förlängning tilldelas även den sista stavelsen i orden *Petronius* och *stoikerna* eftersom prepositionsfraserna *av Petronius* och *för stoikerna* består av minst 4 stavelser.

4.2.1 Val av typ av baskurva

Eftersom ett skifte av baskurva förekommer i meningen består meningen av två typer av baskurvor. Den första baskurvan är av typen ob1a och den andra ob2b. Alla ord i en baskurva märks med samma typ.

4.3 Tilldelning av grundtonsfrekvensnivåer och prominens (RsF0)

Varje stavelse tilldelas en grundtonsfrekvensnivå beroende på om ordet är accent 1, accent 2 eller obetonat. Vilken grundtonsfrekvensnivå som placeras ut i en stavelse beror på vilken

betoningsgrad stavelsen har.

Varje grundtonsfrekvensnivå tilldelas ett prominensvärde (RsF0) som används i beräkningen av grundtonsfrekvensen.

Ef - ter näs - tan sex år i Grek - land for
H* L Hg1 H* L Hg1 L* Ha L* Ha H L* Ha H L* Ha
15 15 15 15 15 15 21 21 20,5 20,5 20,5 20,5 20,5 18,5 18,5 18,5

An - ja - lis till Rom, in - bju - den av Pe - tro - ni - us
L* Ha Lu Lu H L* Ha H* L Hg1 Lu Lu H L* Ha Lu Lu
20,5 20,5 11 11 20,5 20,5 20,5 21,5 21,5 21,5 11 11 20,5 20,5 20,5 11 11

att fö - re - lä - sa för sto - i - ker - na vid fo - rum.
Lu H* L Lu Hg1 Lu H L* Ha Lu Lu Lu Lu H* L Hg1
11 18,5 18,5 11 18,5 11 20,5 20,5 20,5 11 11 11 11 20,5 20,5 20,5

4.4 Tilldelning av prominens (RsD) och beräkning av duration

Durationen av fonem beräknas utifrån fasta värden i millisekunder då fonemet står i betonad och obetonad ställning. I beräkningen ingår även ett prominensvärde för fonemet (RsD). Värdet är detsamma som för ordets prominens i samtliga fall utom då grundtonsfrekvensnivån är av typen Lu, H och Hg1. För Lu och H är prominensvärdet 11. Däremot om Lu direkt efterföljer Hg1 eller Ha i samma ord beräknas prominensvärdet till värdet på föregående Hg1 eller Ha - 5. För beräkning av RsD för Hg1 används formeln $(0,5 * \text{prominens (RsD) för föregående H*}) + 6,1$.

e f t e r n ä s t a n s e k s å r i
15 15 13,6 13,6 13,6 15 15 15 13,6 13,6 13,6 21 21 21 21 20,5 20,5 11

g r e k l a n d f o r a n j a l i s
20,5 20,5 20,5 20,5 11 11 11 11 18,5 18,5 18,5 20,5 20,5 15,5 15,5 11 11 11

t i l r o m i n b j u d e n a v
11 11 11 20,5 20,5 20,5 21,5 21,5 16,85 16,85 16,85 11,85 11,85 11,85 11 11

p e t r o n i u s a t f ö r e l ä s a
11 11 20,5 20,5 20,5 15,5 15,5 11 11 11 11 18,5 18,5 11 11 15,35 15,35 10,35 10,35

f ö r s t o i k e r n a v i d f o r u m
11 11 11 20,5 20,5 20,5 15,5 11 11 11 11 11 11 11 11 20,5 20,5 16,35 16,35 16,35

Fonemdurationerna beräknas i millisekunder. Se 3.4 *Beräkning av duration* för formel och *Appendix* för durationsvärden för fonemen i betonad respektive obetonad stavelse.

e f t e r n ä s t a n s e k s å r i g r e k l a n d
65 99 86 57 42 58 65 99 86 76 66 113 99 153 127 142 59 58 61 56 142 98 37 66 53 76

f o r a n j a l i s t i l r o m i n b j u d e n
105 125 54 101 102 46 83 37 58 81 83 43 36 56 96 102 101 107 55 49 110 46 48 57

a v p e t r o n i u s a t f ö r e l ä s a
76 36 83 43 94 56 142 59 68 43 81 66 75 105 125 37 43 46 97 77 64

f ö r s t o i k e r n a v i d f o r u m
79 58 36 112 94 142 68 75 43 36 54 66 37 58 45 112 142 49 73 80

4.4.1 Final förlängning

De ord som tidigare tilldelats final förlängning är *Rom*, *Petronius*, *stoikerna* och *forum*. Fonemen i ordens sista stavelser förlängs med 60% eller 30% beroende på dess placering och betoning. Den sista stavelsens fonem i ordet *forum* förlängs med 30% eftersom stavelsen är betonad och meningsfinal. Fonemet /r/ förlängs med 15 millisekunder, /u/ med 22 och /m/ med 24. Resterande ord förlängs med 60%. Förlängningen i millisekunder för fonemen /r/, /o/, /m/, /u/, /s/, /n/ och /a/ är 34, 57, 61, 26, 49, 32 och 40.

4.4.2 Klusterreduktion

Vid särskilda konsonantföljder inom samma stavelse eller över stavelsegränserna sker en reduktion av fonemens durationer. I exemplet förekommer klusterreduktion hos orden *sex*, *Grekland* och *stoikerna*.

Två efterföljande konsonanter, förutom /r/, /j/ och /h/ i samma stavelse förkortas med 30%. För ordet *sex* innebär det att fonemet /k/ reduceras med 46 millisekunder och får en duration på 107 millisekunder istället för 153. /s/ reduceras med 38 millisekunder och får en duration på 88 millisekunder. Samma regel gäller för fonemen /n/ och /d/ i ordet *Grekland* och fonemen reduceras med 16 respektive 23 millisekunder.

Fonemen /s/ och /t/ reduceras i ordet *stoikerna* med 30% respektive 20% då det är ett initialt /s/ som efterföljs av /t/ i samma stavelse. Durationen reduceras med 34 respektive 19 millisekunder.

4.5 Tilldelning av relativ position för grundtonsfrekvensnivåer

Samtliga grundtonsfrekvensnivåer i en baskurva summeras för att beräkna dess relativa position. Antalet grundtonsfrekvensnivåer i den första baskurvan är 23 och i den andra 26. Den relativa positionen för grundtonsfrekvensnivåerna beräknas genom att dividera ordningstalet (1, 2, 3 osv.) med summan av antalet grundtonsfrekvensnivåer i baskurvan.

ef	–	ter	...	rom		in	–	bj	–	den	...
H*	L	Hg1	...	L* Ha		H*	L	Hg1	Lu	...	
1	2	3	...	22 23		1	2	3	4	...	
0,0435	0,087	0,1304	...	0,9565 1		0,0385	0,0769	0,1154	0,1538	...	

4.6 Beräkning av grundtonsfrekvens

Beräkningen sker i två steg. Först räknas ett värde ut för grundtonsfrekvensnivån beroende på dess relativa position och typ av baskurva.

Formel för ob1a:

$$-12,6 * relativ\ position^4 + 44,04 * relativ\ position^3 - 52,16 * relativ\ position^2 + 17,6 * relativ\ position + 1,62$$

Formel för ob2b:

$$-37,63 * \text{relativ position}^4 + 64,91 * \text{relativ position}^3 - 40,175 * \text{relativ position}^2 + 7,75 * \text{relativ position} + 1,15$$

ef	-	ter	...	rom		in	-	bju	-	den	...
H*	L	Hg1	L*	Ha		H*	L	Hg1		Lu	...
2,3	2,8	3,1	-1,3	-1,5		1,4	1,5	1,6		1,6	

För beräkning av grundtonsfrekvensen för en obetonad stavelse (Lu) används endast formeln för baskurvan (se ovan). För beräkning av grundtonsfrekvensen för alla andra nivåer beräknas dessutom en formel för varje grundtonsfrekvensnivå. Formeln är densamma för en grundtonsfrekvensnivå i en ob1a, ob1b och ob3a och ytterligare en formel tillämpas för grundtonsfrekvensnivåer i ob2a, ob2b och ob3b.

Formel för H* i ob1a:

$$-0,0212 * RsF0^2 + 1,231 * RsF0 - 11,804 + \text{beräknat värde från formel för ob1a}$$

Formel för L i ob1a:

$$-0,0045 * RsF0^3 + 0,3068 * RsF0^2 - 6,584 * RsF0 + 41,46 + \text{beräknat värde från formel för ob1a} - 22,6 * \text{relativ position}^3 + 50,5 * \text{relativ position}^2 - 31,2 * \text{relativ position} + 4,77$$

Formel för Hg1 i ob1a:

$$-0,003 * RsF0^3 + 0,216 * RsF0^2 - 4,32 * RsF0 + 24,79 + \text{beräknat värde från formel för ob1a}$$

Formel för H* i ob2b:

$$-0,0141 * RsF0^2 + 0,85 * RsF0 - 7,7 + \text{beräknat värde från formel för ob2b}$$

Formel för L i ob2b:

$$-0,0042 * RsF0^3 + 0,2888 * RsF0^2 - 6,217 * RsF0 + 39,4 + \text{beräknat värde från formel för ob2b}$$

Formel för Hg1 i ob2b:

$$0,0456 * RsF0^2 - 1,465 * RsF0 + 11,25 + 28,05 * \text{relativ position}^3 - 35,22 * \text{relativ position}^2 + 8,4 * \text{relativ position} + 1 + \text{beräknat värde från formel för ob2b}$$

Då den obetonade stavelsen (Lu) i *inbjuden* direkt följer grundtonsfrekvensnivån Hg1 är verdet för den obetonade stavelsen grundtonsfrekvensvärdet från Hg1.

ef	-	ter	...	rom		in	-	bju	-	den	...
H*	L	Hg1	L*	Ha		H*	L	Hg1		Lu	...
4,2	1,8	1,6	0,0	1,0		5,4	-1,0	4,0		4,0	

4.7 Utdata

Textens fonetiska skrift, duration, var i fonemet grundtonsfrekvensen ska uppnås och grundtonsfrekvensen skrivs till en textfil. Textfilen är anpassad för indata i talsyntes med MBROLA. Grundtonsfrekvensen räknas om från semitoner till Hertz.

<u>fonetisk skrift</u>	<u>duration</u>	<u>procent</u>	<u>Hertz</u>
[65	0	127
F	99	86	111
T	86		
E0	57		
R	42	100	110
...
R	73		
O	125	0	100
M	132	100	106
-	450	0	100
I	101	0	137
N	107	46	95
B	55		
J	49		
U:	110	100	126
D	46		
[48	50	126
N	57		

Efter nästan sex år i Grekland for Anjalis till Rom,

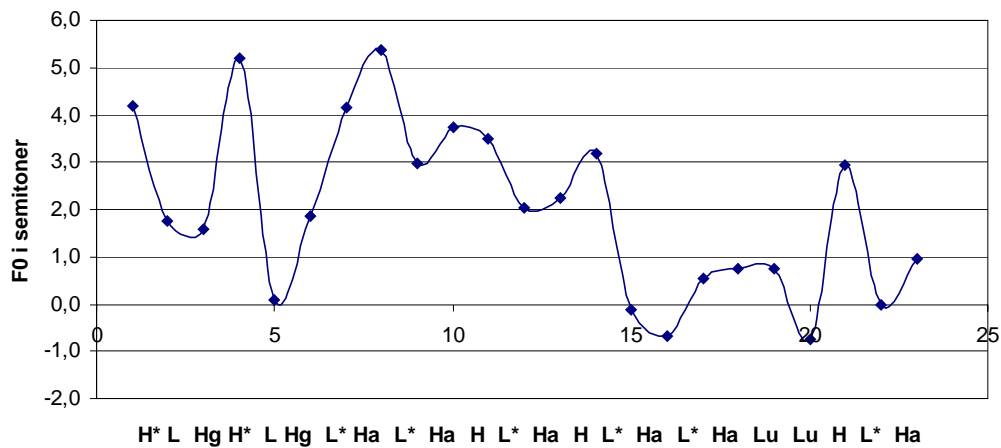


Bild 4.1: Grundtonsförlopp över den första delen av meningen *Efter nästan sex år i Grekland for Anjalis till Rom,*

inbjuden av Petronius att föreläsa för stoikerna vid forum.

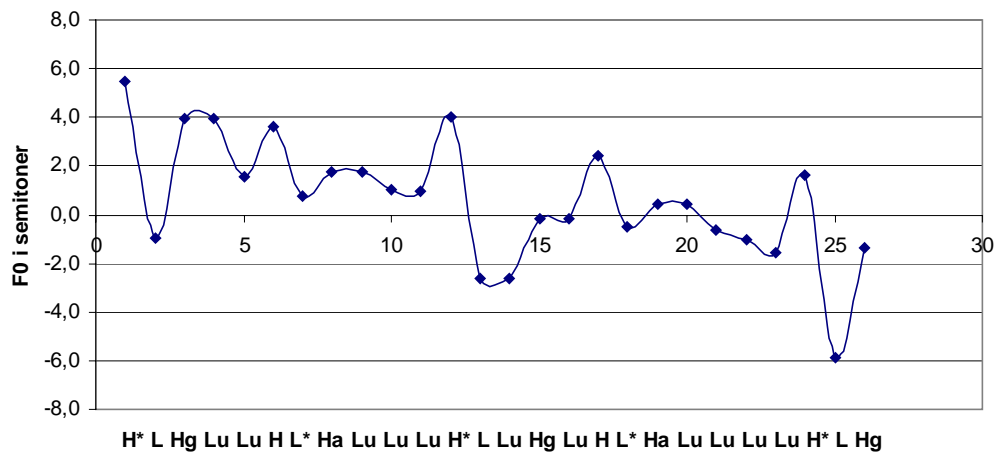


Bild 4.2: Grundtonsförlopp över den andra delen av meningen *inbjuden av Petronius att föreläsa för stoikerna vid forum.*

5 Slutsats

I examensarbetet har en analys av FK-systemet gjorts. Systemets ingående delar har identifierats och strukturerats i pseudokod. FK-systemet är ett omfattande system som hanterar svensk prosodi på detaljnivå. En implementering i pseudokod utifrån systemets definierade regler visade sig dock vara möjlig med undantag från de förutsättningar som modellen kräver. Då FK-systemet utvecklas kontinuerligt har reglerna arbetats om efter att nyligen genomförda experiment och revideringar har skett. Det innebär att senare versioner av modellen finns än den som är implementerad här. Implementeringen är överskådlig och modulär, vilket kan förenkla tillägg och förändringar i koden allteftersom nya regler utformas. Även om implementeringen är modulär kräver FK-systemet samtliga beräkningar och tilldelningar av värden för att utdata ska kunna genereras.

Förutsättningarna modellen kräver gäller främst bearbetningen av texten samt innehållet i lexikon. De måste uppfyllas för att en implementering ska vara möjlig. Därmed ställs höga krav på modellen redan initialt. Den parser som används måste kunna identifiera en menings samtliga huvudsatser och ingående bisatser samt fraser. I de lexikon som används måste ett ords stavelser och dess betoningsgrad finnas med samt vilka ingående ord en sammansättning fogats samman av. I dagsläget är modellens utdata anpassat till en fonemorienterad talsyntes.

Resultatet av implementeringen leder till en tänkbar prosodisk beskrivning av en text. Grupperingen av yttrandet är i huvudsak beroende av indelningen av intonationskurvor och bygger enbart på syntax. FK-systemet har endast applicerats på ett par texter ur romaner och dagstidningar med en relativt grammatiskt korrekt uppbyggnad, vilket kan vara en bidragande faktor till att en verklighetstrogen talsyntes kan genereras för den typen av texter. Eftersom språk är föränderligt kan materialet till indelningen av intonationskurvor tyckas vara något snäv.

Prosodin hjälper oss att semantiskt tolka ett yttrande. Det är därför av stor vikt att kunna identifiera de semantiskt betydelsefulla delarna i en text redan på ett tidigt stadium. Tolkningen av en text kan skilja sig åt beroende på dess semantiska uppbyggnad. Exempelvis spelar kontextuella förhållanden en avgörande roll där *ny* respektive *given* information ges. Kontexten är dessutom betydelsefull för att kunna generera prosodi på grammatiskt inkorrekta texter. Det är därför önskvärt att systemet ska kunna göra en kontextuell analys av texten.

5.1 Framtida utveckling

Ännu behandlar FK-systemet endast ett urval av fraser och ordklasser. En noggrann lingvistisk analys av alla de fraser och ordklasser som ska hanteras är nödvändig för att modellen ska bli komplett. Idag hanteras exempelvis endast nominal- och prepositionsfraser. Det vore önskvärt att även adverb- och adjektivfraser hanterades.

Prediceringen av fokus i FK-systemet är påbörjad. I nuläget krävs vidare forskning för att kunna skriva regler för prediceringen av fokala ord. Ett fokalt ords prominensvärde höjs markant vilket föranleder en förhöjd grundtonsfrekvens. Även durationen påverkas och måste beräknas på nytt. I ett fokalt ord förlängs den betonade stavelsen samt den postfokala stavelsen. Graden av förlängning beror bland annat på redan befintliga delar i modellen, exempelvis huruvida ordet är accent 1 eller 2 samt ordets position i satsen. För att kunna identifiera ett fokalt ord behövs främst en utveckling av en kontextuell analys av texten där ordens semantik och främst diskurs spelar en avgörande roll för om ett ord är fokalt eller inte.

Då ett längre yttrande avslutas sker ett finalt fall, sk. terminal junktur. Accentmoduleringen i modellen följer Bruces (1977) benämningar och har implementerats därefter. Däremot saknas en implementering av terminal junktur. För att markera terminal junktur i FK-systemet påbörjas fallet i mitten av den betonade stavelsens vokal på intonationsmodulens sista ord. En extra lågpunkt läggs till sist i intonationskurvan för att generera en markant sänkning i grundtonsfrekvens och ett naturligt avslut på den prosodiskt samhöriga gruppen.

En intressant vidareutveckling av implementeringen av FK-systemet är att implementera modellen så att ingångsvärdena som krävs för beräkningar av grundtonsfrekvens och duration matas in manuellt. På så sätt skulle vidare forskning på prosodimodellen kunna förenklas och sambandet mellan grammatiska- samt syntaktiska enheter och prosodi utrönas. Därefter och först då systemets förutsättningar implementerats vore en implementering i ett befintligt text-till-talsystem tänkbar.

Bibliografi

Barbosa Ferreira, J. (2003). *Pauser och jukturer i uppläst tal*. Examensuppsats, Högskolan i Skövde.

Bhaskararao, P. (1994). Subphonemic segment inventories for concatenative speech synthesis. I Keller, E. (red.), *Fundamentals of Speech Synthesis and Speech Recognition: Basic concepts, State of the Art, and Future Challenges*. Chichester: John Wiley. S. 69-85.

Bruce, G. (1977). *Swedish word accents in sentence perspective*. Diss., Lunds universitet. Lund: LiberLäromedel.

Bruce, G. (1998). *Allmän och svensk prosodi*. Lund: Institutionen för lingvistik, Lunds universitet. (Praktisk lingvistik, 16).

Carlsson, R. & Granström, B. (1973). *Word accent, emphatic stress and syntax in a synthesis by rule scheme for Swedish*. Speech Transmission Laboratory QPSR 2-3, Kungliga Tekniska Högskolan, Stockholm. S. 31-36.

Ceder, K. & Lyberg, B. (1992). *Yet another rule compiler for text-to-speech conversion?* Proceedings ICSLP 92, 12-16 oktober, Banff, Kanada. S. 1151-1154.

Dutoit, Thierry (1996). *A Short Introduction to Text-to-Speech Synthesis*. [Elektronisk]. Faculté Polytechnique De Mons, Belgien. Tillgänglig: <http://tcts.fpms.ac.be/synthesis/mbrola.html> [2004-08-27].

Dutoit, T., Pagel, V., Pierret, N., Bataille, F. & van der Vrecken, O. (1996). *The MBROLA Project: Towards a set of high quality speech synthesizers free of use for non commercial purposes*. Proceedings ICSLP 96, oktober, Philadelphia, U.S.A. S. 1393-1396.

Fant, G. (2001). *Huvuddragen i Prediktion*. Opublicerat manuskript.

Fant, G. (2002). *Prosodiregler system FK*. Opublicerat manuskript.

Fant, G. (2003a). *Preliminära regler för jukturer och pauser*. Opublicerat manuskript.

Fant, G. (2003b). *Sammanfattning avseende Ob kurvor och pauser*. Opublicerat manuskript.

Fant, G. & Kruckenberg, A. (1999). *Syllable and word prominence in Swedish*. Stockholm: Kungliga Tekniska Högskolan.

Fant, G. & Kruckenberg, A. (2001). *Prosodiregler i sammandrag*. Opublicerat manuskript.

Fant, G., Kruckenberg, A. & Barbosa Ferreira, J. (2003). *Individual variations in pausing – A study of read Speech*. Stockholm: Kungliga Tekniska Högskolan.

Fant, G., Kruckenberg, A. & Gustafson, K. & Liljencrants, J. (2002). *A New Approach to Intonational Analysis and Synthesis of Swedish*. TMH-QPSR Vol. 44, Kungliga Tekniska Högskolan, Stockholm. S. 161-164.

- Fant, G., Kruckenberg, A. & Liljencrants, J. (2000). Acoustic-phonetic Analysis of Prominence in Swedish. I Botinis (red.), *Intonation: Analysis, Modelling and Technology*. Dordrecht: Kluwer Academic Publishers (Text, Speech and Language Technology series. Vol. 15).
- Fredriksson, M. (1988). *Den som vandrar om natten...* Trondheim: Aktietrykkeriet i Trondhjem.
- Frid, J. (1999). *An environment for testing prosodic and phonetic transcriptions*. ICPH 99, 1-7 augusti, San Francisco, U.S.A. S. 2319-2322.
- Frid, J. (2003). *Lexical and Acoustic Modelling of Swedish Prosody*. Diss., Lunds universitet. Lund:Studentlitteratur.
- Heldner, M. (1998). Is an F0 Rise a necessary or a sufficient cue to perceived Focus in Swedish? I Werner, S. (red.), *Nordic Prosody*. Frankfurt am Main: Europäischer Verlag der Wissenschaften. S.109-125.
- Heldner, M. & Strangert, E. (2001). Temporal Effects of Focus in Swedish. *Journal of Phonetics*. Vol. 29, s. 329-361.
- Horne, M. & Filipsson, M. (1996). Computational extraction of lexico-grammatical information for generation of Swedish intonation. I van Santen, J. P. H., Sproat, R. W., Olive, J. P. & Hirschberg, J. (red.), *Progress in Speech Synthesis*. New York: Springer Verlag. S. 443-457.
- Ladd, R., D. (2000). Tones and Turning Points: Bruce, Pierrehumbert, and the Elements of Intonational Phonology. I Horne, M. (red.), *Prosody: Theory and Experiment. Studies presented to Gösta Bruce*. Dordrecht: Kluwer Academic Publishers (Text, Speech and Language Technology series. Vol. 14).
- Lindblom, B., Lyberg, B. & Holmgren, K. (1976). *Durational Patterns of Swedish Phonology: Do They Reflect Short-term Memory Process?* Bloomington: Indiana university.
- Lyberg, B. (1981). *Temporal Properties of Spoken Swedish*. Diss., Stockholms universitet. Edsbruk: Holms Gårds Tryckeri.
- Monaghan, Alex (1992). *Generating Synthetic Prosody: Means & Ends*. [Elektronisk]. University of Edinburgh, Skottland. Tillgänglig: <http://www.compapp.dcu.ie/~alex/PUB/aix92.html> [2004-08-27].
- O'Shagughnessy, D. (2000). *Speech Communications, Human and Machine*. New York: IEEE Press.
- Portele, T. & Heuft, B. (1996). Towards a Prominence-based synthesis system. *Speech Communication*. Vol. 21, s. 61-72.
- Werner, S. & Keller, E. (1994). Prosodic aspects of speech. I Keller, E. (red.), *Fundamentals of Speech Synthesis and Speech Recognition: Basic concepts, State of the Art, and Future Challenges*. Chichester: John Wiley. S. 23-40.

Appendix

Durationsvärden för fonem i betonad (Do) respektive obetonad (Db) stavelse.

Före vokal:

	Do	Db
/k/, /p/, /t/	83	93
/g/, /b/, /d/	45	60
/m/, /n/, /ng/	54	64
/h/, /j/, /l/, /r/, /v/	37	55
/s/, /sj/, /tj/, /f/	79	110

Efter kort vokal:

	Do	Db
/k/, /p/, /t/	75	145
/g/, /b/, /d/	76	110
/m/, /n/, /ng/	53	99
/h/, /j/, /l/, /r/, /v/	36	80
/s/, /sj/, /tj/, /f/	81	122

Efter lång vokal:

	Do	Db
/k/, /p/, /t/	75	97
/g/, /b/, /d/	66	76
/m/, /n/, /ng/	53	74
/h/, /j/, /l/, /r/, /v/	36	58
/s/, /sj/, /tj/, /f/	81	127

Korta vokaler:

	Do	Db
/a/	66	99
övriga	43	93

Långa vokaler:

	Do	Db
/a/	76	158
övriga	58	138