Samenvatting

Intonatie De intonatie van een uiting omvat alle aspecten van het waargenomen toonhoogtepatroon die de luisteraar gebruikt om die uiting te begrijpen, of die de spreker daartoe heeft aangebracht. In het algemeen worden drie aspecten van intonatie onderscheiden: zinsaccenten, intonationele grenzen, en zinsmelodie. Een *zinsaccent* verleent een woord nadruk, en correspondeert met een relatief snelle verandering van de toonhoogte rondom de beklemtoonde lettergreep van het woord. Een *intonationele grens* klinkt als een lichte onderbreking in het spreken (hoewel er niet noodzakelijk een pauze hoeft te zijn) die wordt waargenomen als gevolg van een verandering in de toonhoogte rondom de grens. De *zinsmelodie* betreft de specifieke vorm van het toonhoogtepatroon, die onafhankelijk van de plaats van accenten en grenzen kan variëren. Zo kan een mededeling worden omgezet in een vraag zonder de accenten of de grenzen te verplaatsen, maar enkel door de melodie te veranderen.

Net zoals Nederlandse woorden en zinnen *structureel welgevormd* moeten zijn, d.w.z. moeten gehoorzamen aan de grammatica van het Nederlands, zo moet de intonatie van een uiting minimaal voldoen aan de intonationele grammatica. Daarnaast moet de intonatie ook passen bij de vorm, betekenis en context van een uiting. Sprekers van het Nederlands kunnen beoordelen of de intonatie van een gegeven uiting acceptabel is in een bepaalde context. Vanuit het oogpunt van de generatieve taalkunde is het doel om deze oordelen te kunnen voorspellen en verklaren. Hiertoe wordt een taalkundig model opgesteld in de vorm van een formele grammatica die alle acceptabele intonatiepatronen kan genereren, maar geen onacceptabele.

Hierbij doet zich een aantal methodologische problemen voor. Acceptabiliteitsoordelen kunnen namelijk variatie vertonen doordat ze gradueel van aard zijn (de gekozen intonatie kan in meer of mindere mate acceptabel zijn), doordat verschillende alternatieven even acceptabel kunnen zijn, of doordat proefpersonen onderling enigszins verschillen in hun oordelen. Verder is het moeilijk om perceptie-experimenten op te zetten waarbij ongeoefende luisteraars een zuiver oordeel over de intonatie kunnen vellen, aangezien ze de benodigde vaardigheid in analytische luisteren missen. Daarom is er in dit proefschrift voor gekozen om een taalkundig model te maken voor de intonatie die het *meest acceptabel* is (waarbij verschillen tussen even acceptabele varianten worden genegeerd) voor *deskundige luisteraars* (intonationele experts).

Spraakgenerering De centrale vraag binnen het onderzoeksgebied van de *natuurlijke taalgenering* is hoe we computers zo kunnen programmeren dat ze een interne representatie van informatie (bijvoorbeeld een tabel met treinreisgegevens) kunnen omzetten in een goed geschreven tekst (bijvoorbeeld een gemakkelijk te begrijpen treinreisbeschrijving). Een verwant vakgebied, dat van de *spraaksynthese*, beoogt computers tekst (of een meer abstracte fonologische representatie) om te laten zetten in goed verstaanbare en prettig klinkende spraak. *Spraakgenering* is een uitbreiding op taalgenerering zodanig dat de genenereerde tekst vervolgens zo goed mogelijk wordt uitgesproken. Een spraakgenereringssyteem is daarom te zien als een combinatie van een taalgeneringssyteem en spraaksynthese systeem.

Taal- en spraakgenerering zijn multidisciplinaire vakgebieden, waarin de redenen voor het bouwen van generingssystemen kunnen verschillen. Zo is men vanuit de psycholinguïstiek geïnteresseerd in het modelleren van menselijke taalproductie, terwijl men vanuit de taal- en spraaktechnologie streeft naar praktisch inzetbare systemen. De invalshoek in dit proefschrift is primair taalkundig: het spraakgeneringssysteem wordt gebruikt als een hulpmiddel voor het toetsen en verbeteren van taalkundige beschrijvingen van intonatie.

Bestaande taalkundige modellen voor intonatie zijn doorgaans slechts in beperkte mate getoetst. Het probleem is dat ze kennis vereisen over de syntax (de bouw) en semantiek (de betekenis), en de discourse (de context, inclusief de kennis van de spreker en luisteraar) van de uiting die als invoer dient. Veel van deze informatie is niet automatisch uit een zin of tekst af te leiden, en moet daarom met de hand worden gespecificeerd. Dit is tevens één van de redenen dat taalkundige modellen niet direct geschikt zijn voor toepassing in praktische tekst-naar-spraaksystemen. Echter, binnen een spraakgenereringssysteem kan al de benodigde informatie automatisch worden gegenereerd door de taalgenereringsmodule (mits deze voldoende geavanceerd is). Zodoende kunnen dus grote hoeveelheden betrouwbaar testmateriaal worden geproduceerd. Verder kan de daadwerkelijke implementatie van een taalkundig model verborgen aannames, mogelijke tegenspraken en andere gebreken aan het licht brengen.

Onderzoeksdoel Het centrale onderzoeksdoel in dit proefschrift is het ontwerpen, bouwen en testen van een taalkundig gemotiveerd spraakgenereringsysteem dat zinsaccenten en intonationele grenzen genereert die het meest acceptabel zijn volgens intonationele experts. Hierbij gelden een aantal restricties. Ten eerste betreft het voorgelezen spraak (in tegenstelling tot spontane spraak), en is de doeltaal het Nederlands. Ten tweede is het inherent aan de huidige taalgeneringsystemen dat ze gebonden zijn aan een bepaald domein. Dit onderzoek betreft genereringsystemen voor twee typen tekst: treinsreisbeschrijvingen, en botanische beschrijvingen van bloemen. Ten derde concentreert dit onderzoek zich op het genereren van een fonologische representatie van de intonatie, waarbij wordt geabstraheerd van de fonetische implementatie (hoewel wel degelijk hoorbare spraak wordt gesynthetiseerd). Tenslotte is er om praktische redenen voor gekozen om één bepaalde taalkundige beschrijving van de Nederlandse intonatie te evalueren.

Het onderzoek is gestuurd door enerzijds onderzoeksvragen betreffende de taalkundige theorie:

1. Hoe adequaat is de theorie?
2. Hoe kan deze verder worden verbeterd?

Anderzijds, door onderzoeksvragen betreffende de vereiste architectuur van spraakgeneringssystemen:

1. Wat is de beste manier om taalgenerering en spraaksynthese te combineren?
2. Wat is de beste plaats binnen een spraakgenereringssysteem om de intonatie te beregelen?
3. Welke aanpassingen zijn vereist aan de 'standaard' architecturen voor taalgeneringssystemen en spraaksynthese systemen?
4. Wat voor extra informatie moet worden gegenereerd door het taalgenereringssysteem, en hoe kan dit worden bereikt?

Theoretisch kader Hoofdstuk 2 beschrijft het theoretische kader van dit onderzoek. Allereerst wordt het begrip *prosodie* gedefinieerd en wordt de *prosodische fonologie* geïntroduceerd, een fonologische theorie over de hiërarchische organisatie van spraak. Volgens deze theorie heeft iedere spraakuiting een *prosodische structuur*, die samen met de tonale structuur het prosodische deel vormt van de fonologische representatie. Deze representatie wordt opgebouwd op basis van de semantiek, de syntaxis en de morfologie. Er wordt een kort overzicht gegeven van de welgevormdheidscondities voor de prosodische structuur en de argumenten die het bestaan ervan ondersteunen. Eén van de constituenten is de *fonologische frase*, die een belangrijke rol speelt in de rest van dit proefschrift.

Ten tweede introduceert dit hoofdstuk de *intonationele fonologie*, die de contrastieve intonatiepatronen van de taal definieert, meer in het bijzonder de autosegmentele intonationele fonologie van het Nederlands ontwikkeld door Gussenhoven. In navolging van het werk van Pierrehumbert, kent dit model een *toonstring* en een *segmentstring*, die onderling gesynchroniseerd zijn door middel van *associaties*. De toonstring bestaat uit toonhoogteaccenten (H*L, L*H, of H*LH), grenstonen (%H, %L, H%, or L%), en toonmorfemen zoals *downstep* (!). De segmentstring is bovendien opgedeeld in *associatie domeinen* (AD's), die in de praktijk overeenkomen met intonationele frasen. Er worden een aantal voordelen van dit model aangevoerd met betrekking tot het genereren van intonatie: (1) het biedt een precieze formalisering van accentplaatsing, intonationele frasering, en melodiekeuze; (2) het staat een concentratie op de fonologische beschrijving toe, doordat geabstraheerd wordt van de fonetische implementatie; (2) het voorkomt het genereren van intonatie die onacceptabel is op grond van fonologische onwelgevormdheid.

Ten derde wordt in navolging van Gussenhoven beargumenteerd dat de gangbare opvatting dat de intonationele frase kan worden geanalyseerd als een prosodische constituent tot tegenspraak leidt, en dat een onafhankelijk domein is vereist (het associatiedomein).

Ten vierde wordt in dit hoofdstuk een eenvoudige, regelgebaseerde grammatica voor het plaatsen van zinsaccenten beschreven, die het uitgangspunt vormt voor het verdere werk. De invoer bestaat uit de focusdistributie, de argumentstructuur, de prosodische structuur, en de opeenvolgende woorden. De eerste regel is SAAR (Sentence Accent Assignment Rule), die de focus-accent relatie modelleert, in het bijzonder voor gefocusseerde maar ongeaccentueerde predicaten. De tweede regel is de *ritmeregel*, een fonologische regel die onder bepaalde condities accenten deleert om een meer ritmische verdeling daarvan te bewerkstelligen. Er wordt beargumenteerd dat aangezien het toepassingsdomein van deze regel de fonologische frase is, het voor accentplaatsing van cruciaal belang is dat de fonologische frasering correct verloopt.

Tenslotte beschrijft dit hoofdstuk de minimale aannames over de syntactische structuur van het Nederlands die in dit onderzoek worden gehanteerd.

Fonologische frasering Hoofdstuk 3 betreft algoritmes voor fonologische frasering, die binnen het huidige kader essentieel zijn als toepassingsdomein voor de ritme regel. In het eerste deel van dit hoofdstuk worden twee algoritmes beschreven zoals voorgesteld door respectievelijk Nespor & Vogel en door Hayes, en wordt beargumenteerd dat geen van beide adequaat is voor het Nederlands. Vervolgens wordt een nieuw algoritme voorgesteld dat berust op de notie van *adjacente modificatie*. Naar aanleiding van het testen op een groot aantal verschillende syntactische structuren van het Nederlands volgt een gedeeltelijke herziening van het algoritme. Hoewel verdere generalisatie wenselijk is, voorziet het resultaat in een observationeel adequaat algoritme voor het Nederlands.

Het tweede deel van dit hoofdstuk beschrijft een evaluatie van dit algoritme door te controleren of toepassing van de ritmeregel binnen de voorspelde fonologische frasen tot de gewenste accentuering leidde. Deze evaluatie werd uitgevoerd op symbolisch niveau, met een gedeeltelijk automatische, corpusgebaseerde methode. Het corpus bestond uit 10 korte bloembeschrijvingen (935 woorden), die handmatig werden geparseerd, en door een drietal intonationele experts werden voorzien van de geprefereerde accenten. De voorspellingen van het accentueringsalgoritme, inclusief het fonologische fraseringsalgoritme en de ritmeregel, werden vervolgens vergeleken met de referentie. Voorspelling en referentie kwamen in 95% van alle gevallen overeen. Een analyse van dat deel van de fouten dat mogelijk te wijten zou zijn aan verkeerde toepassing van de ritmeregel, gaf geen directe aanleiding om het algoritme voor fonologische frasering te herzien.

Intonationele frasering Het eerste deel van Hoofdstuk 4 beschrijft een drietal algoritmes voor intonationele frasering zoals voorgesteld door respectievelijk Nespor & Vogel, Caspers, en Gussenhoven. Vervolgens wordt een evaluatie van deze algoritmes beschreven die analoog is aan de evaluatie beschreven in het vorige hoofdstuk, namelijk, op symbolisch niveau, en met behulp van van een gedeeltelijk automatische, corpusgebaseerde methode. Hiervoor werd het eerder gebruikte corpus van bloemenbeschrijvingen door de intonationele experts eveneens voorzien van de meest acceptabele intonationele grenzen. Daardoor konden de voorspellingen van de drie algoritmes worden vergeleken met deze referentie. Er wordt een gedetailleerde analyse van de fouten beschreven, die laat zien dat elk van de drie algoritmes systematische tekortkomingen heeft. Er wordt betoogd dat een beter algoritme de syntactische structuur als invoer zou moeten gebruiken, en daarbij gevoelig zou moeten zijn voor bepaalde syntactische configuraties, de accentdistributie, en het gewicht van de gevormde intonationele frasen.

In het tweede deel wordt een algoritme voorgesteld dat aan deze voorwaarden voldoet. Het bestaat uit acht regels die intonationele grenzen inserteren in bepaalde syntactische configuraties, mits aan een minimale verdeling van de accenten is voldaan. Een evaluatie van dit algoritme op het corpus van bloembeschrijvingen liet zien dat het in 97.66% van alle gevallen een correcte beslissing nam. Een statistische test toonde aan dat het algoritme significant beter was dan tenminste twee van de andere algoritmes.

Spraakgenering Hoofdstuk 5 gaat in op de onderzoeksvragen betreffende de architectuur van een spraakgeneringssysteem. De voorgestelde architectuur is een uitbreiding op de conventionele architectuur voor taalgenereringsystemen bestaande uit een pijplijn van drie modules. Voor wat betreft het combineren van taalgenering met spraaksynthese wordt beargumenteerd dat een verbinding via tekst ontoereikend is. In plaats van tekst wordt een fonologische representatie voorgesteld die bestaat uit de woorden, de intonationele vorm (accenten, grenzen, en melodie), en de hogere prosodische constituenten (waaronder de fonologische frase). Ook wordt betoogd dat prosodische realisatie (dat ondermeer het genereren van de intonationele vorm omvat) niet kan worden gereduceerd tot microplanning of oppervlakte realisatie, maar een complex, en inherent linguïstische proces is, dat een eigen module in het systeem vereist.

Vervolgens wordt een architectuur uitgewerkt bestaande uit twee hoofddelen. Ten eerste een domeinspecifieke tekstplanner die de *inhoud* van de tekst genereert in de vorm van een reeks abstracte specificaties voor syntactische frasen. Ten tweede een domeinonafhankelijk en herbruikbaar *inhoud-naar-spraaksysteem* (CONPAS genaamd) dat verantwoordelijk is voor de *vorm* van de gesproken uitvoer, gebruik makend van morfo-syntactische, fonologische, en fonetische kennis van Nederlands. Dit deel bestaat intern weer uit een drietal modules voor syntactische realisatie (SynR), prosodische realisatie (ProsR), en spraaksynthese (SpeechS). Verder wordt beargumenteerd dat de gebruikte specificaties voor syntactische frasen voorzien in het juiste niveau van abstractie.

De rest van dit hoofdstuk biedt een overzicht over de taken die binnen de verschillende modules worden uitgevoerd. Ten eerste worden de relatief eenvoudige, schemagebaseerde textplanners beschreven, die primair tot doel hebben zoveel mogelijk variatie in tekst te genereren binnen de twee toepassingsdomeinen (treinreisbeschrijvingen en bloembeschrijvingen). Ten tweede worden syntactische en prosodische realisatie kort samengevat; beide processen worden uitgebreid beschreven in de volgende twee hoofdstukken. Tenslotte worden de drie stappen in spraaksynthese beschreven: fonologie (afleiding van de lagere prosodische constituenten, en toepassing van segmentele regels), fonetische implementatie (beregeling van duur, pauze, en F_0), en difoonsynthese.

Syntactische realisatie Hoofdstuk 6 beschrijft SynR, een syntactische realisator voor het Nederlands die abstracte specificaties vertaalt naar grammaticale expressies en hun syntactische structuur, zodat een gebruiker/systeem kan abstraheren van de syntactische details van de doeltaal. Deze uitwerking laat zien dat alle syntactische informatie die nodig is voor het plaatsen van accenten en intonationele grenzen (zoals verondersteld in de eerdere hoofdstukken) inderdaad automatisch kan worden gegenereerd door een taalgeneringsssysteem.

Het systeem bestaat uit drie opeenvolgende componenten. Ten eerste een basis component, die hoofdzakelijk verantwoordelijk is voor het projecteren van de basale syntactische structuur, het toevoegen van puur syntactische informatie, en het uitschrijven van getallen, tijden, etc. Ten tweede een transformationele component die zorgt voor ondermeer de distributie van focus, verschillende vormen van congruentie, en de woordvolgorde. Ten derde een flectie component die inhoudswoorden vervoegt en functiewoorden invoegt. Het systeem voorziet in een onderscheid tussen de declaratieve/linguïstische en de procedurele/computationele aspecten, doordat alle modules zijn geïmplementeerd in speciale grammaticaformalismen (de transformationele grammatica in Gramtsy, en de andere twee als de affix grammatica's in AGFL).

Er wordt betoogd dat de voorgestelde abstracte invoer een compacte specificatie toelaat, dankzij de notationele conventies en de aanname van redelijke defaults. Verder wordt beargumenteerd dat een systeem als SynR een aantal voordelen biedt boven een eenvoudige, templaat-gebaseerde syntactische realisator, ondermeer voor wat betreft prosodische realisatie en microplanning. Tenslotte wordt SynR kort vergeleken met andere systemen, en worden een aantal mogelijke uitbreidingen besproken.

Prosodische realisatie Hoofdstuk 8 beschrijft ProsR, de module voor de prosodische realisatie van uitingen op basis van hun syntactische structuur, focusdistributie, argumentstructuur, en woordenreeks. Dit proces omvat een viertal taken: (1) het afleiden van de hogere prosodische constituenten (waaronder de fonologische frase); (2) het plaatsen van accenten; (3) het plaatsen van intonationele grenzen; (4) het kiezen van de melodie. Aangezien de eerste en derde taak reeds zijn behandeld in respectievelijk hoofdstuk 3 en 4, wordt in dit hoofdstuk hoofdzakelijk aandacht besteed aan accentplaatsing.

Op basis van de eenvoudige grammatica uit hoofdstuk 2 wordt een meer adequate methode voor accentplaatsing voorgesteld. Allereerst worden de details van de implementatie van SAAR uitgewerkt, ondermeer met betrekking tot scheidbare werkwoorden, en meer in het algemeen, samenstellingen die meer dan één accent kunnen dragen. Verder worden een aantal uitzonderingen op SAAR besproken. Er wordt betoogd dat focus projectieregels noodzakelijk zijn om de relatie tussen focus op woordniveau en op constituentniveau te definiëren. Ook wordt een aantal resterende problemen bediscussieerd, waaronder gevallen van interactie tussen SAAR en intonationele frasering.

Vervolgens worden een aantal extra regels geïntroduceerd en gemotiveerd. Ten eerste een regel voor het deaccentueren van naamwoordgroepen die fungeren als voorbepalingen bij een zelfstandig naamwoord. Ten tweede een lexicale ritmeregel voor een beperkt aantal lexicaal gespecificeerde quantificeerders en bijwoorden die ritmisch gezien een uitzondering vormen. Ten derde een speciale ritme regel voor eenheden (waaronder *honderd* en *duizend*), die in combinatie met de juiste fonologische frasen de accentuering van getallen correct voorspelt.

Het laatste deel van dit hoofdstuk presenteert een eenvoudig algoritme voor melodiekeuze, waarin ondermeer een alternatieve vorm van downstep wordt gesignaleerd.

Evaluatie Hoofdstuk 8 beschrijft een evaluatie-experiment waarbij de accenten en intonationele grenzen geproduceerd door het generingsysteem werden vergeleken, op symbolische niveau, met die voorgeschreven door een viertal intonationele experts. Tevens vond een vergelijking plaats met een representatief tekst-naar-spraaksysteem. Het materiaal bestond uit acht reisbeschrijvingen (643 woorden) en 60 expressies met getallen (274 woorden). De resultaten lieten zien dat voor de reisbeschrijvingen de accentuering van het systeem in 93% van alle gevallen overeenkwam met die van de experts, terwijl dit voor de getalsexpressies 95% bedroeg. Voor wat betreft de intonationele frasering werd in beide domeinen een overeenstemming van 95% gevonden.

Vervolgens worden deze resulaten bediscussieerd. Er wordt een hiërarchische clusteranalyse gepresenteerd die inzicht verschaft in de onderlinge verschillen tussen de systemen en de experts. Verder worden waar mogelijk systematische fouten geanalyseerd.

Tenslotte worden de resultaten vergeleken met die voor andere systemen. Hier worden ondermeer metrieken voor het berekenen van de score bediscussieerd. Er wordt geconcludeerd dat de accentuering van het ontwikkelde spraakgeneringssysteem beter is dan dat van de meeste tekst-naar-spraaksystemen, en vergelijkbaar met die van andere spraakgenereringssystemen, terwijl de frasering beter is dan die van andere systemen in termen van overeenstemming, en vergelijkbaar in termen van andere metrieken.

Algemene discussie Hoofdstuk 9 vat het proefschrift samen aan de hand van de antwoorden op de onderzoeksvragen betreffende de architectuur van spraakgeneringssystemen en de linguïstische theorie over intonatie. Ter afsluiting wordt kort aandacht besteed aan de praktische implicaties.

Curriculum Vitae

Erwin Marsi was born in Deventer on November 27 1968. He attended the Eemland College in Amersfoort, where he obtained his VWO diploma in 1986. He started his study at Nijmegen University with Philosophy, but changed after one year to Language, Speech and Computer Science. He enjoyed additional courses in Cognitive Science, Philosophy, Computer Science, and Mathematics, and graduated cum laude in 1993. From 1994 to 1998, he was employed as a PhD student (AIO) at the Department of Language and Speech, Nijmegen University. While continuing work on his thesis, he was part-time employed at Nijmegen University for teaching a course on Natural Language Generation. He recently joined the Department of Computational Linguistics and AI at Tilburg University as a postdoc researcher, and collaborates in a project about pitch accent placement and prosodic phrasing using techniques from machine learning and information retrieval.