

## Summary

In this thesis we have investigated several aspects of the spectro-temporal structure of vowel segments, both concerning vowel production as well as vowel perception. Chapter 1 contains a summary of current models on vowel production and perception. Models of vowel pronunciation try to explain why vowel realizations vary so much in natural speech. It is known that vowel production is influenced in highly systematic ways by context, stress, and speaking style (among others). The classical explanation is that of the target-undershoot model. This model states that vowel articulation is limited by the speed of the articulators (e.g., jaw, tongue, lips). Each vowel has a unique target-position for each of the articulators which will produce the ideal, or canonical, realization of that vowel. When vowel realizations are very long, there is ample time for the different articulators to reach their respective target positions. However, when vowel duration is short and the context forces the articulators to cover relatively large distances, there is not enough time and the articulators are stopped short of their targets. The resulting vowel realizations show "undershoot" in their articulatory movements as well as in the resulting formant frequencies, hence the name of the model: target-undershoot.

The classical quantitative study of Lindblom (1963) on the relation between vowel duration and formant-undershoot is discussed in depth. It showed that formant-undershoot increased exponentially with a decrease in vowel duration. However, subsequent studies gave ambiguous results. Some studies did find clear evidence for articulatory- and formant-undershoot. Others showed that there were numerous cases where no relation between vowel duration and target-undershoot could be found. Especially, changes in stress and speaking style could bring about changes in duration that were not accompanied by changes in target-undershoot. In our opinion, these conflicting results can be explained by assuming that target-undershoot is planned by the speaker. In this view, the undershoot serves a purpose that depends on factors like context, prosody, and speaking style. From this it follows that, irrespective of vowel duration, the undershoot itself should not change if the purpose of the undershoot does not change and vice versa.

Considering the conflicting reports in the literature, it seems that any test of the target-undershoot model should introduce changes in vowel duration without changing stress, speaking style, or other prosodic factors that were known to cause ambiguous results. In this study, we settled for changes in speaking rate. A long, meaningful text, read at a normal and at a fast rate, would induce a speaker to use the same stress assignments and the same "style" of speaking, irrespective of reading speed. At the same time, a difference in speaking rate would change the duration of all the vowels. In this study (chapters 2-4), we used all realizations of seven different vowels and some realizations of the schwa (/ʌ/). If vowel duration could control formant-undershoot all by itself, then an increase in speaking rate should induce an increase in undershoot. However, if formant-undershoot is planned, then a change in speaking rate should not necessarily result in a change in formant-undershoot.

In chapter 2, we measured formant frequencies in the *vowel kernel*. Vowel realizations uttered at the normal speaking rate were compared to the corresponding realizations uttered at the fast speaking rate. No spectral vowel reduction was found that could be attributed to a faster speaking rate. There was also no change in the amount of coarticulation or stress-induced reduction as a result of speaking rate. The only systematic effect was a higher  $F_1$  value in fast-rate speech irrespective of vowel identity. This possibly suggests a generally more open articulation of vowels, speaking louder, or some other general change in speaking style by our speaker when he speaks fast.

In chapter 3 we looked at the effects of speaking rate on *vowel formant track shape*, using the same material as in chapter 2. The formant track shape was assessed on a point-by-point basis, using 16 samples at the same relative positions in the vowels. Differences in speaking rate only resulted in the same uniform change in  $F_1$  frequency already found in chapter 2. Within each speaking rate, there was only evidence for a weak leveling off of the  $F_1$  tracks of the open vowels /A a/ with shorter durations. When considering sentence stress or vowel realizations from a more uniform, alveolar-vowel-alveolar context, these same conclusions were reached.

In chapter 4 we again looked at the effects of speaking rate on formant track shape. This time we used a more elaborate method for assessing formant track shape. Legendre polynomial functions were used to model and quantify the shape of time normalized formant tracks. No differences in these normalized formant track shapes were found either that could be attributed to differences in speaking rate. A uniform higher  $F_1$  frequency in fast-rate speech relative to normal-rate speech was found. Within each speaking rate, there was only evidence of a weak leveling off of the  $F_1$  tracks of the open vowels /E A a/ with shorter durations. Again, as in chapter 3, separately inspecting vowel realizations from a more uniform, alveolar-vowel-alveolar context, did not alter our conclusions.

The target-undershoot model of vowel production inspired a complementary model of vowel perception (Lindblom and Studdert-Kennedy, 1967). As vowel formant tracks will systematically undershoot the canonical target values in natural speech, it was suggested that listeners would compensate for this undershoot automatically by systematically overshooting the formant frequencies actually reached in perception, i.e. perceptual-overshoot. Early studies with synthetic speech did indeed find this kind of perceptual-overshoot. However, it showed to be rather difficult to prove the existence of an automatic mechanism for perceptual-overshoot in natural speech.

At the moment, there are two classes of models on vowel perception. The first class are models with dynamic-specification. In these models it is assumed that listeners use dynamical information from the Consonant-Vowel and/or Vowel-Consonant transitions to improve the recognition of the, stationary, vowel nucleus. Perceptual-overshoot is just one of such models. The second class of models is based on the assumption that a single, spectral, cross-section of the kernel of a vowel realization contains all information necessary to recognize it. In these models the vowel on- and offset transitions are of minor importance in vowel recognition.

The difference between these two types of models is the position of the Consonant-Vowel transition (in the vowel on- and offset). Is it used in vowel recognition, as is stated by models using dynamic-specification, or is it not, as stated by target models? There is evidence for perceptual-overshoot in synthetic speech. It is also known that presenting syllables without a vowel kernel, i.e. with only the vowel on- and offset transitions, hardly impairs vowel recognition. Still, there is no undisputable proof that the recognition of isolated, monphthongal, vowel segments is improved by adding dynamical information to the formant tracks. Exactly such an improvement is expected when listeners use dynamic-specification of vowels.

In natural speech, the amount of variation in durations, vowel formant frequencies and track shapes is limited. These various types of variation are furthermore strongly correlated. It is therefore better to use synthetic speech, for which it is possible to control all features. With synthetic speech, it is also possible to detach formant track shape from formant frequency. This way, the effects of formant track shape can be studied independently of vowel identity and vowel duration. We therefore choose to use synthetic speech to study how vowel duration and formant track shape influence vowel identity. Especially we looked for any evidence for perceptual-overshoot. The result of this study is presented in chapter 5 (see below). In chapter 6 we took a closer look at the existing literature in order to try to find an explanation for the disagreement between our results and those presented in several earlier papers.

In chapter 5 we used synthetic vowels to investigate whether listeners use vowel duration and formant track shape to determine vowel identity. The synthetic vowels had level or parabolically-shaped formant tracks and variable durations. They were presented in isolation as well as in synthetic CVC syllables. There was no evidence of perceptual compensation for expected target-undershoot due to token duration or context. The only asserted effects of duration and context were in the number of long- and short-vowel responses. There was also no evidence that the listeners used the formant track shape or slopes independently to identify the synthetic vowel tokens. Tokens with curved formant tracks were generally identified near their formant offset frequencies.

The results of chapter 5 contradicted claims made in the literature about the way listeners use dynamical information to identify vowel realizations. The literature on vowel perception itself also contains contradictory claims regarding the use of information from CV-transitions in vowel recognition. Our own experiments showed that the information in formant track shape was not always used to compensate for formant-undershoot. In chapter 6 a re-evaluation of the literature is attempted. A closer study of the most relevant papers shows that evidence for compensatory processes, i.e. perceptual-overshoot and dynamic-specification, was only found when vowel realizations from different, and appropriate, context were contrasted. Some studies show that vowel recognition deteriorated when vowel segments were presented out of context. Together, these facts suggest that the presence of an appropriate context is essential for any perceptual compensation of coarticulatory changes. This speculation might be used as a starting hypothesis for further research on vowel perception.

Finally, in chapter 7 we summarize and discuss our findings. We recapitulate the methods used in chapters 2 to 4 to study the effects of speaking rate on formant-undershoot. We argue that, under the circumstances used, any excess undershoot due to an increase in speaking rate should have been detectable, but did not show up. We therefore conclude that, for our speaker, speaking rate did not influence the amount of vowel formant-undershoot or the formant track shape. Therefore, we can conclude that changes in vowel duration alone do not change the amount of target-undershoot and that the undershoot that does occur is probably planned.

The listening experiments presented in chapter 5 showed that our listeners did not use a perceptual-overshoot mechanism or dynamic-specification to help them identifying the synthetic vowel tokens. In general, they seemed to use the offset part of each vowel realization to identify it. We therefore conclude that listeners do not automatically and unconditionally compensate for the formant-undershoot that can be predicted from the formant track shape.

## Samenvatting

Beschrijving en identificatie van klinkers lijkt een simpel probleem te zijn. Wanneer klinkers echter door machines herkend moeten worden, of omgekeerd, wanneer machines klinkers moeten produceren, dan wordt de complexiteit van dit probleem al snel duidelijk. Klinkers zoals *aa*, *ie* of *oe* kunnen articulatorisch beschreven worden met slechts drie parameters: 1) de mate waarin de mond open is; 2) de positie van de tong, voor/boven of achter/beneden; 3) de mate waarin de lippen getuit zijn. Bij de klinker *aa*, zoals in *vaas*, is de mond zo ver mogelijk open, ligt de tong "middenin" de mond en zijn de lippen gespreid. Bij de klinker *ie*, zoals in *fiets*, is de mond (bijna) gesloten, ligt de tong vóór in de mond en (bijna) tegen het verhemelte en zijn de lippen gespreid. Bij de klinker *oe*, zoals in *voet*, is de mond ook gesloten, maar ligt de tong achterin de mond en zo ver mogelijk van het verhemelte en zijn de lippen getuit. Deze drie klinkers zijn het meest extreem wat betreft de positie van onderkaak, tong en lippen (de articulatoren). De andere Nederlandse klinkers liggen ertussenin.

Wanneer het geluid van klinkers onderzocht wordt lijkt de zaak in eerste instantie zelfs nog simpeler te worden. Klinkers worden onderscheiden op hun klankkleur (naar analogie van het timbre van muziekinstrumenten). De klankkleur van een klinker kan grotendeels beschreven worden met slechts twee frequenties, die van de eerste twee resonanties van de mondkeelholte. Deze resonanties worden formanten genoemd ( $F_1$  en  $F_2$ ). De klinkers *ie* en *oe* hebben de laagste waarde voor de  $F_1$  en respectievelijk de hoogste en de laagste waarde voor de  $F_2$ . De klinker *aa* heeft de hoogste waarde voor de  $F_1$  en een gemiddelde waarde voor de  $F_2$ . Wanneer de frequentie van de tweede formant uitgezet wordt tegen de frequentie van de eerste formant dan vormen de *ie*, *oe* en *aa* de hoekpunten van een driehoek. De waarden voor de formanten van alle andere klinkers (b.v. *uu*, *oo*, *o*, *ee*, *e*, *eu*) liggen binnen deze klinkerdriehoek.

Voor langgerekte klinkers, zoals 'aaaaaaah' of 'oooooooooh' geldt nu een heel simpele regel: bij elke klinker hoort een unieke waarde voor de  $F_1$  en  $F_2$ . Als men weet wat voor klinker uitgesproken is, dan weet men ook vrij nauwkeurig wat de waarden van de eerste twee formanten zullen zijn. Omgekeerd, als men de waarden van de eerste twee formanten kent, dan weet men ook wat voor klinker er uitgesproken is.

Helaas is het in werkelijkheid niet allemaal zo eenvoudig. Omdat de formanten resonanties zijn van de mond-keelholte, zijn ze afhankelijk van de grootte van mond en keel. Dit wil zeggen dat de frequenties van deze formanten anders zullen zijn voor mannen, vrouwen en kinderen. En ook binnen deze groepen zijn de individuele verschillen groot. Deze variatie kan berekend worden en men kan ervoor corrigeren. Na correctie, of wanneer men de spraak van één enkele spreker bekijkt, geldt de eenvoudige, één-éénduidige relatie tussen langgerekte klinkers en formantfrequenties weer.

Nu is het verleidelijk om deze eenvoudige relatie tussen formantwaarden en (langgerekte) klinkers door te trekken naar normale spraak. Dit blijkt echter niet zomaar te kunnen. Er zijn verschillende processen die roet in het eten gooien. Allereerst is er een proces dat *coarticulatie* genoemd wordt. Als je goed luistert, dan hoor je dat de *a* uit *kar* niet hetzelfde klinkt als die

uit *tas*. De formantwaarden die men kan meten voor deze twee realisaties van de *a* zijn ook duidelijk verschillend. Het lijkt in deze gevallen of de medeklinkers die om de klinker staan, de formantwaarden ervan in de richting van een zeer specifieke frequentie 'trekken'. Als men alle mogelijke combinaties van klinkers en medeklinkers onderzoekt, dan blijkt dat er een grote spreiding bestaat in de formantwaarden van dezelfde klinkers. Het komt relatief vaak voor dat de ene klinker in de ene context dezelfde formantwaarden heeft als een andere klinker in een andere context. Zonder de context te kennen is het vaak niet meer mogelijk om te voorspellen wat de formantwaarden zullen zijn van een klinker en andersom is het niet meer mogelijk om uit enkel de formantwaarden te bepalen om welke klinker het gaat.

Er is nog een tweede proces dat de klankkleur en daarmee de formantwaarden van klinkers verandert. Dit proces wordt *reductie* genoemd. In dezelfde omgeving van medeklinkers, klinkt de *a* uit *kaboutter* toch anders dan die uit *kabbelen*. Er is ook een verschil in formantwaarden. Het verschil tussen *kaboutter* en *kabbelen* is te wijten aan woordklemtoon. In *kaboutter* zit de *a* in een onbeklemtoonde lettergreep, in *kabbelen* in een beklemtoonde. Naast woordklemtoon en zinsaccent speelt ook de stijl van spreken een rol. Als iemand een tekst voorleest dan praat hij/zij anders dan wanneer hij/zij een ongedwongen gesprek voert. Gemiddeld genomen liggen de formantwaarden van onbeklemtoonde klinkers en klinkers uit ongedwongen conversatie meer in het midden van de klinkerdriehoek dan de beklemtoonde klinkers en de klinkers uit voorgelezen teksten. Het lijkt erop alsof de formantfrequenties gemiddeld naar het centrum van de klinkerdriehoek getrokken worden. Onbeklemtoonde klinkers en klinkers uit ongedwongen conversatie zijn *gereduceerd* ten opzichte van beklemtoonde klinkers en klinkers uit voorgelezen tekst.

Coarticulatie en reductie zijn twee verschijnselen die de klankkleur van klinkers sterk en systematisch veranderen. Als gevolg hiervan is het niet meer mogelijk om op grond van alléén de formantfrequenties de identiteit van de klinker te achterhalen (machinaal of automatisch klinkers herkennen is moeilijk). Toch blijkt dat menselijke luisteraars er weinig moeite mee hebben om klinkers in welke context dan ook te herkennen. Met betrekking tot klinkers zijn er nu twee vragen waarop een antwoord gezocht wordt. Ten eerste, hoe verandert de klankkleur van klinkers als gevolg van context, klemtoon en spreekstijl? Met andere woorden, welk mechanisme zit er achter coarticulatie en reductie? Ten tweede, hoe zijn luisteraars in staat een klinker te herkennen ondanks het feit dat de klankkleur door coarticulatie en reductie sterk verandert? Over deze twee vragen gaat dit proefschrift.

In hoofdstuk 1 van dit proefschrift wordt een overzicht gegeven van de relevante literatuur. We bespreken o.a. de klassieke studie van Lindblom uit 1963. In deze studie vindt Lindblom dat er een relatie is tussen de duur van een klinker en de formantwaarden. Lindblom formuleerde een model waarbij de duur van een klinker de mate van coarticulatie, en indirect die van reductie, bepaalde. Dit model wordt het 'target-undershoot' model genoemd ('het doel niet bereiken'; er is geen goede Nederlandse vertaling voor deze term). Dit model gaat uit van het feit dat de menselijke articulatoren,

zoals onderkaak, tong en lippen, tijd nodig hebben om de bewegingen te maken die nodig zijn om medeklinker-klinker-medeklinker reeksen uit te spreken. De snelheid waarmee deze organen bewogen kunnen worden is beperkt. Als er te weinig tijd is, kunnen de noodzakelijke bewegingen niet meer afgemaakt worden, het doel wordt gemist, en er ontstaat coarticulatie. Lindblom beweerde nu dat de duren van de klinkers in normale spraak eigenlijk al te kort zijn om nog perfect uitgesproken te kunnen worden. Later werd dit model genuanceerd door te stellen dat de mate waarin de klinkerduren te kort zijn, afhankelijk is van de inspanning die de spreker zich getroost om de klinkers goed uit te spreken. Ook bij deze nuancering blijft echter gelden dat de duur van de klinkers de mate van coarticulatie en reductie bepaalt.

Sinds het target-undershoot model werd geformuleerd zijn er diverse studies uitgevoerd waarvan de resultaten dit model ondersteunden, maar ook studies die het model tegenspraken. Het bleek bijvoorbeeld, dat onbeklemtoonde klinkers best even lang kunnen zijn als beklemtoonde klinkers, terwijl ze toch gereduceerd zijn. Het kwam ook voor dat klinkers wel korter werden, maar zonder dat er meer coarticulatie of reductie optrad. Het lijkt zeer wel mogelijk dat zowel coarticulatie als reductie (ten dele) 'bewust' uitgevoerd worden, en dat de relatie tussen coarticulatie, reductie en klinkerduur ontstaat doordat de duur van een klinker meestal ook korter wordt in omstandigheden die leiden tot coarticulatie en reductie. Dit betekent dat de articulatoren wel degelijk sneller kunnen bewegen dan dat ze normaal doen en dat het mogelijk moet zijn om spraak uit te lokken met veel kortere klinkers maar zonder extra coarticulatie en reductie. Om nu de geldigheid van het target-undershoot model te onderzoeken moet de klinkerduur variëren terwijl alle andere factoren die kunnen leiden tot verschillen in coarticulatie en reductie (zoals klemtoon, context, spreekstijl e.d.) hetzelfde blijven. Een van de manieren waarop dit bereikt kan worden is door een spreker te vragen een tekst voor te lezen, eerst in een normaal tempo, daarna zo snel mogelijk. Dit is de methode die wij voor ons onderzoek gekozen hebben.

In de hoofdstukken 2 tot en met 4 hebben wij onderzocht of er inderdaad meer coarticulatie en reductie optreedt wanneer klinkers korter worden in snelle spraak. We gebruikten daarvoor een normale tekst die twee keer werd voorgelezen door een professionele nieuwslezer, eerst in een normaal tempo, daarna snel. We gebruikten zeven van de twaalf Nederlandse klinkers (de *oe*, *oo*, *a*, *aa*, *e*, *ie* en *uu*) en verder enkele realisaties van de schwa (de *uh* klinker uit 't en d'r). De klinkers waren zo gekozen dat ze goed verspreid lagen over de 'klinkerdriehoek'. In hoofdstuk 2 onderzoeken we de formantwaarden in het midden van iedere klinkerrealisatie. Het bleek dat er geen noemenswaardig verschil was tussen de formantwaarden in snelle en normale klinkers. Voor alle klinkers was er een lichte stijging in de frequentie van de eerste formant die misschien het gevolg is geweest van een verschil in luidheid. Het kan zijn dat onze spreker niet alleen sneller maar ook harder is gaan praten. Niets wijst er echter op dat er ook maar enig verschil in coarticulatie of reductie is tussen de normaal en de snel gelezen versie van de tekst.

Aangezien de formantwaarden gerelateerd zijn aan de positities van de articulatoren, kunnen verschillen in de bewegingen (meestal) teruggevonden worden door de formantwaarden te volgen in de tijd. Als de vorm van de formantsporen (d.w.z. de sporen van de formantfrequenties in de tijd) verschilt tussen normale en snelle spraak, dan moeten ook de bewegingen van de articulatoren verschillen. Als de vorm van de formantsporen, na normalisatie voor duur, niet verschilt tussen normale en snelle spraak, dan is het onwaarschijnlijk dat de bewegingen van de articulatoren wel verschillen.

In de hoofdstukken 3 en 4 gebruiken we twee verschillende methoden om vormverschillen in formantsporen te onderzoeken, na eerst voor de klinkerduur gecorrigeerd te hebben. In hoofdstuk 3 gebruiken we een rechttoe-rechtaan methode om te onderzoeken of er verschillen zijn tussen begin, midden en eind van iedere klinker. In hoofdstuk 4 gebruiken we een meer geavanceerde methode (hogere orde curve fitting) om te kijken of de formantsporen vlakker worden in snelle spraak. Geen van beide methoden toont enig verschil tussen normale en snelle spraak aan. Hieruit moet geconcludeerd worden dat er wel een verschil is in de *snelheid* van de bewegingen van de articulatoren in normale en snelle spraak van onze professionele spreker, maar geen verschil in het *verloop* van de bewegingen.

In hoofdstuk 7 lichten we nogmaals toe dat het target-undershoot model een meetbare toename van coarticulatie en reductie zou hebben voorspeld in snelle spraak. Wij vinden echter geen verschil. Hieruit moet geconcludeerd worden dat onze spreker in staat is sneller te spreken zonder *extra* coarticulatie en reductie (d.w.z. boven de normale variatie) en dat deze verschijnselen dus niet direct afhangen van de klinkerduur. Hieruit volgt dat het waarschijnlijker is dat coarticulatie en reductie actief geregeld worden door onze spreker en dat zij niet het gevolg zijn van de passieve traagheid van zijn articulatoren.

In het perceptieve deel van dit proefschrift (hoofdstukken 5 en 6) onderzoeken we hoe luisteraars klinkers identificeren. Uit het voorgaande moge gebleken zijn dat klinkers nogal variëren wat betreft klankkleur. De vraag is nu hoe luisteraars toch in staat zijn deze variabele klanken te identificeren. In hoofdstuk 1 zijn de twee belangrijkste theoriën op dit punt besproken. De eerste theorie is op het eerste gezicht de eenvoudigste. Deze stelt dat de formanten in het midden van de klinker genoeg informatie bevatten om de klinker te identificeren. Men kan er echter niet meer mee volstaan om de eerste twee formanten te gebruiken ( $F_1$  en  $F_2$ ), maar men moet ook de derde formant ( $F_3$ ) en de grondtoon, d.w.z. de toonhoogte ( $F_0$ ) gebruiken. Tevens is het noodzakelijk om formanten aan elkaar te relateren (bijvoorbeeld,  $F_3$ - $F_2$  i.p.v. de afzonderlijke formanten).

De tweede theorie stelt dat de problemen ontstaan door de gevolgen van coarticulatie. Als bekend is hoe de coarticulatie 'gericht' is, dan kunnen de gevolgen ongedaan worden gemaakt. De richting en mate van coarticulatie kunnen bepaald worden door de formantsporen aan het begin en einde van een klinker te bekijken, daar waar klinker en medeklinker elkaar raken. Door de hellingen van de formantsporen aan het begin en eind te extrapoleren kan men een goede schatting maken van de 'doelwaarde', d.w.z. de



waarde zonder coarticulatie. Het hypothetische proces in het menselijke gehoor dat daarvoor moet zorgen wordt 'perceptual-overshoot' genoemd ('waarnemingscorrectie door extrapolatie', deze term is ook al niet goed te vertalen). Het gehoor trekt, als het ware, de bewegingen van de formanten, en daarmee die van de articulatoren, door. Het principe dat de vorm van de formantsporen in het overgangsgebied tussen naburige medeklinkers en klinkers een rol speelt bij de identificatie wordt 'dynamische specificatie' genoemd. Er zijn veel artikelen geschreven over de voors en tegens van deze twee theoriën over klinkeridentificatie, maar tot nog toe spreken de resultaten elkaar tegen.

In hoofdstuk 5 onderzoeken wij de effecten van de vorm van formantsporen op de identificatie van klinkers. Dit is gedaan door synthetische klinkers aan luisteraars aan te bieden. In deze synthetische klinkers werden de duur van de klinker en de vorm van de formantsporen (hele en halve parabolen) gevarieerd. Wij kunnen duidelijk aantonen dat onze luisteraars de formantsporen in onze synthetische klinkers niet extrapoleren naar een hypothetische waarde zonder coarticulatie. Integendeel, in plaats van de bewegingen in de klinkeraanzet te extrapoleren wordt over het laatste deel van de klinker gemiddeld.

In hoofdstuk 6 gaan we dieper in op de tegenstrijdige resultaten in de literatuur, alsook in die van ons eigen onderzoek. In experimenten van anderen waarin aanwijzingen gezocht worden voor het bestaan van perceptual-overshoot en dynamische specificatie, worden de reacties van luisteraars onderzocht op klinkers met verschillende typen formantsporen. Het blijkt echter dat de vorm van de formantsporen niet de enige factor is die gevarieerd werd in die experimenten. Telkens wanneer aanwijzingen gevonden worden voor perceptual-overshoot en dynamische specificatie blijkt ook de context gevarieerd te zijn. Die experimenten zijn op een dusdanige manier uitgevoerd dat compensatie voor de coarticulatie als gevolg van de context en extrapolatie van formantsporen precies hetzelfde resultaat zouden hebben gehad. Het blijkt ook dat het weglaten van de context de verstaanbaarheid van klinkers zeer nadelig beïnvloed. In hoofdstuk 7 concluderen we uit de resultaten van ons eigen onderzoek (hoofdstuk 5) dat perceptual-overshoot en dynamische specificatie niet zonder meer volgen uit de vorm van de formantsporen. Samen met de resultaten van anderen, zoals gedetailleerd beschreven in hoofdstuk 6, suggereert dit dat een 'passende' context noodzakelijk is voor compensatie van de effecten van coarticulatie. Het kan zelfs zijn dat de context op zichzelf al voldoende is om een luisteraar aan te zetten tot het compenseren van eventuele effecten van coarticulatie.