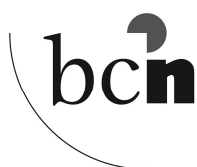


# Prosodic Processes in Language and Music

Maartje Schreuder

Copyright © 2006 by Maartje Schreuder  
Cover design: Hanna van der Haar  
Printed by Print Partners Ipskamp, Enschede



The work in this thesis has been carried out under the auspices of the  
Research School of Behavioral and Cognitive Neurosciences (BCN),  
Groningen



Groningen Dissertations in Linguistics 60  
ISSN 0928-0030  
ISBN 90-367-2637-9

RIJKSUNIVERSITEIT GRONINGEN

# Prosodic Processes in Language and Music

Proefschrift

ter verkrijging van het doctoraat in de  
Letteren  
aan de Rijksuniversiteit Groningen  
op gezag van de  
Rector Magnificus, dr. F. Zwarts,  
in het openbaar te verdedigen op  
donderdag 15 juni 2006  
om 13:15 uur

door

Maartje Johanneke Schreuder

geboren op 25 augustus 1974  
te Groningen

Promotor: Prof. dr. J. Koster

Copromotor: Dr. D.G. Gilbers

Beoordelingscommissie: Prof. dr. J. Hoeksema  
Prof. dr. C. Gussenhoven  
Prof. dr. P. Hagoort

## Preface

Many people have helped me to finish this thesis. First of all I am indebted to my supervisor Dicky Gilbers. Throughout this dissertation, I speak of 'we'. That is not because I have some double personality which allows me to do all the work in collaboration, but because Dicky was so enthusiastic about the project that we did all the experiments together. The main chapters are based on papers we wrote together for conference proceedings, books, and journals. This collaboration with Dicky always was very motivating and pleasant. I will never forget the conferences we visited together, especially the fun we had trying to find our way in Vienna, and through the subterranean corridors in the castle in Imatra, Finland. Music has always been the central theme in our conversations, though we rarely agreed about the question what real music is. Well, we learnt to appreciate each others musical choices. I persuaded him to open his ears to classical music with real cellos, and he, for his part, convinced me that the 60s and 70s produced many beautiful songs. The differences in our musical appreciation have found their way into the various musical examples in this dissertation. For all that, and for the time he invested in supervising me, I owe him a lot. So Dicky: thank you!

I owe a debt also to my promotor Jan Koster for his valuable comments on my work. Jack Hoeksema, Carlos Gussenhoven, and Peter Hagoort kindly agreed to be on my reading committee, for which I would like to thank them. This thesis benefited particularly from the comments of Carlos Gussenhoven, and from our useful discussion on the recursion data. I also thank John Nerbonne for reading it thoroughly, as the director of the CLCG. For corrections to the musicological part of this thesis, I express my gratitude to Paul van Reijen, who prevented me from making too many mistakes in the musical terminology. Obviously, I remain responsible for all remaining shortcomings. I also wish to thank Fred Lerdahl, Grzegorz Dogil, Stephanie Shattuck Hufnagel, Henkjan Honing, and Norman Cook for their useful comments.

I am also very grateful to our trained listeners Dicky Gilbers, Femke Wester, Wander Lowie, Dirk-Bart den Ouden and Jack Hoeksema for doing the time-consuming job of judging our data, to our subjects for participating in our experiments, and to Korine Bolt

and Aline Klingenberg for letting me recruit subjects among the students in their classes. Many more people have helped me with this thesis. I especially thank Paul Boersma, Wilbert Heeringa, Hugo Quené, and Wouter Jansen for supplying us with the PRAAT scripts that we used for the analyses, and Femke Wester, Dieuwke de Goede, Laurie Stowe, John Nerbonne, and Vincent van Heuven for helping with the statistic analyses. Dicky Gilbers and I had a very fruitful collaboration with Tamás Bíró working on a paper on Simulated Annealing and fast speech. This was very useful for both parties: Tamás could use our fast speech data, and we were happy to make use of his analysis. I wish to thank Anthony Runia for reviewing my English, and Rogier Blokland for reviewing the English of earlier versions of some chapters. Again, all remaining mistakes are my own responsibility. Hanna van der Haar designed the wonderful cover of this dissertation, with which I am very pleased. I am most obliged to Laura van Eerten, who as a research assistant annotated the burden of our data on textgrids in PRAAT. For the major/minor experiments she gathered the data, and she is also responsible for part of the analyses. She did a very fine job and I hope she will find a research position in the near future.

Anna Hausdorf, Rob Visser, and secretaries Wyke van de Meer, Tineke Datema, Belinda Houwen, Annemieke Koning, Alja Mensing, Alice Pomstra, Jolanda Westra, M. Slegers, Natalie Specken, Jochum Algra, and Gorus van Oordt ensured things ran smoothly in the department. The experiments could not have been done without the help of Gert Heise and Jan de Ruiter of the audio-visual department. And whenever my computer failed, the helpdesk, especially Vincent Boxelaar, always helped me out.

Working in the Linguistics department has always been a great pleasure. I appreciated the useful discussions with the CLCG Phonology and Phonetics Group: Tjeerd de Graaf, Dicky Gilbers, Toshi Shiraishi, Angela Grimm, Markus Bergmann, Dirk-Bart den Ouden, Wander Lowie, Charlotte Gooskens, Tamás Bíró, Wouter Jansen, and Nanne Streekstra. I also owe Toshi Shiraishi, Tjeerd de Graaf and Markus Bergmann for allowing me to join them in their office, thus saving me from exile in the Pelsterstraat building. Before that, Rogier Blokland, Paul van Linde and Angela Grimm were great office mates for over three years. I thank them especially for the hilarious Friday afternoons. I am grateful to all other colleagues, for

the lunches and drinks together, and for sharing our PhD frustrations, and, of course, for the linguistic discussions: Femke Wester, Jantien Donkers, Julia Klitsch, Liefke Reitsma, Nienke van den Bergh, Katrien Colman, Janneke ter Beek, Dieuwke de Goede, Marjolein Deunk, Rasmus Steinkrauss, Erik-Jan Smits, Holger Hopp, Laura Sabourin, Judith Rispens, Eshter Ruigendijk, Joanneke Prenger, Sible Andringa, Nesli Yetkiner, John Hoeks, Rienk Withaar, Petra Hendriks, Jennifer Spenader, Bart Hollebrandse, Roel Jonkers, Evelien Krikhaar, Jan-Wouter Zwart, Mark de Vries, Elzerieke Hilbrandie, Eleonora Rossi, Maria Trofimova, Tuba Yarbay Duman, Monika Zempléni, Stéphanie Bakker, Joost Zwarts, Tanja Gaustad, Menno van Zaanen, Mark-Jan Nederhof, Begoña Villada Moiron, Leonoor van der Beek, Lonneke van der Plas, Jori Mur, Jörg Tiedemann, Robbert Prins, Gerlof Bouma, Gosse Bouma, Gertjan van Noord, George Welling, Leonie Bosveld, Roelien Bastiaanse, Gerard Bol, Ron van Zonneveld, Gisela Redeker, Ger de Haan, and Frans Zwarts.

I am greatly indebted to my parents and my grandparents for their indispensable support, and I express my gratitude to Harman for providing me with the necessary distraction and support at home and for helping me with some musical tips. Many thanks to my family and friends for showing their interest and for being so understanding and patient. Finally, I thank in advance Marjon Witteveen and Jantien Donkers for standing by me at the public defence of my thesis.





# Contents

<b>Introduction</b>	<b>1</b>
<b>Chapter 1 Language and Music in Optimality Theory</b>	<b>5</b>
1.1. Introduction	5
1.2. The resemblances between language and music	7
1.2.1 Structuring	7
1.2.2 Conflicting preference rules	10
1.2.2.1 <i>Evaluation of possible output candidates</i>	10
1.2.2.2 <i>A linguistic example of conflicting constraints</i>	13
1.2.2.3 <i>A musical example of conflicting constraints</i>	17
1.2.2.4 <i>Boundary marking</i>	23
1.3. Summary and Conclusion	30
<b>Chapter 2 Rhythm</b>	<b>33</b>
2.1. Introduction	33
2.2. Rhythm and meter	35
2.3. Rhythm units	38
2.3.1 Rhythm typology	38
2.3.2 Musical rhythm parallels the linguistic rhythm typology	40
2.4. Variable speech rhythm	41
2.4.1 Eurhythm in speech	41
2.4.2 Triplet rhythm in trochaic Dutch	44
2.5. Rhythmic timing	49
2.5.1 Restructured rhythm in music	49
2.5.2 Timing in speech rhythm	51
2.6. Summary	53
<b>Chapter 3 The Influence of Speech Rate on the Perception of Rhythm Patterns</b>	<b>55</b>
3.1. Introduction	55
3.2. Data	57

3.3.	Framework and phonological analysis	58
3.3.1	Rhythmic restructuring in music	58
3.3.2	Rhythmic restructuring in speech	61
3.3.3	Alternative OT accounts of variation	63
3.4.	Method	72
3.4.1	Subjects and task design	72
3.4.2	Analysis methods	73
3.5.	Results	75
3.5.1	Evaluating the task design	75
3.5.2	Auditory analysis	76
	3.5.2.1 <i>Between-type variation</i>	79
	3.5.2.2 <i>Between-item variation</i>	84
	3.5.2.3 <i>Between-subject variation</i>	86
3.5.3	Phonological analysis: Simulated Annealing	87
3.5.4	Acoustic analysis	91
3.6.	Conclusion	102
<b>Chapter 4</b>	<b>Recursion in Phonology</b>	<b>105</b>
4.1.	Introduction	105
4.2.	Recursion	105
4.2.1	Droste effect	106
4.2.2	Fractals in nature	108
4.2.3	Endless loops in music and visual art	109
4.2.4	Recursively embedded structures in music	111
4.2.5	Computing recursion	116
4.3.	Recursion in Phonology	116
4.3.1	Strict Layering and recursion	116
4.3.2	Research question	120
4.4.	The experiment	122
4.4.1	Task design	122
4.4.2	Subjects	123
4.4.3	Method	124
4.4.4	Data	126
4.4.5	Results	128
	4.4.5.1 <i>Auditory results</i>	128

4.4.5.2 <i>Acoustic results</i>	132
4.4.6 Phonological analysis	142
4.5. Conclusion	148
<b>Chapter 5 Speaking in Minor and Major keys</b>	<b>151</b>
5.1. Introduction	151
5.2. Theoretical background	152
5.3. Method	154
5.4. Analysis and results	155
5.4.1 Cluster analysis	155
5.4.2 Musical scores	159
5.5. Conclusion	163
<b>Chapter 6 Summary, Conclusions, and Future directions</b>	<b>165</b>
6.1. Summary and conclusions	165
6.2. Future directions	168
<b>References</b>	<b>169</b>
<b>Samenvatting</b>	<b>183</b>
<b>Groningen Dissertations in Linguistics</b>	<b>189</b>



## Introduction

Everyone is familiar with the phenomenon of the ticking of a clock. A clock goes *tick tock, tick tock* in English, or *tik tak, tik tak* in Dutch. Why do we assign different vowels to the two ticks of a clock? Do the two ticks sound different? Some clocks have a mechanism that does in fact produce different sounds, but many clocks do not have such a mechanism. Nevertheless, those clocks are imitated as *tick tock, tick tock*. It is our own imagination that assigns a ‘tick’ to the first ‘tick’ and a ‘tock’ to the second. The fact that we do this collectively says something about the way our cognition works.

How does it work? Our cognition wishes to hear structure, in order to understand everything around us in easily manageable chunks. It wants to know the coherence of things. Therefore, everything is accommodated into hierarchies, with several levels, and one element on each level as the most important one, the head element. In the ticking of a clock, we decide that the two ticks form a group of two, and the first element of this group of two ticks is the most important one, to which we ascribe the /ɪ/-sound, which is an intrinsically higher vowel than the /ɔ/-sound. We also think that we perceive it as louder, longer, and higher in pitch.

By accommodating all elements of a sounding object – or visual objects or movements – to a hierarchy of important and less important elements, the interpretational task is made easier. This dissertation is on language and music, which are both cognitive behaviors of people, concerned with sound. Several ways lead to the decision which elements are most important, and which groups of elements form domains together, such as syllables, feet, or phrases. In language and music, domains are based on the cohesion in meaning, structure, or form, or on distance or difference from other elements. Groupings on the basis of meaning (semantics) can differ from those based on phonological structure, which in their turn can differ from groupings on the basis of syntactic structure. Intonation, rhythm, pauses, etc. add their own grouping phenomena. In music similar influences play a role in structuring: melody, rhythm, rests, chord progressions, etc. From all these “cues”, our cognition has to

choose the most straightforward way to assign saliency to some elements, whereas other elements are seen as ornamentation.

In Chapter 1 we will show how these choices are made. We will describe a musical and a linguistic theory, which both give a similar account of how our cognition makes these decisions, defining the preference rules our cognition seems to make use of. We will show that language and music are rather similar in the above-mentioned respects, and in this dissertation we will investigate whether some processes occurring in language, in particular in speech prosody, can be explained on the basis of musical theory.

The main chapters are Chapters 3, 4 and 5, in which we describe the results of three prosody experiments.<sup>1</sup> Chapter 2 provides an introduction to the background theory on rhythm, which can be seen as the most obvious shared characteristic of language and music. The chapter gives an overview of some theoretical issues of the subject.

Chapter 3 concerns an experiment on rhythmic restructuring of secondary stress in words. An important issue in prosodic variability research is, for instance, the question whether the influence of a higher speaking rate leads to adjustment of the phonological structure or just to phonetic compression, or maybe just to a different perception by the listener. On the basis of the similarities between language and music and the insight that restructuring can occur in rate adjustments in music, we suppose that phonological adjustment/restructuring on account of differences in speaking style and speaking rate is possible. For this issue the phonologist could profit from the musicologist's knowledge.

Another subject of considerable debate in linguistics is that a mismatch seems to exist between syntactic structure and phonological structure. Syntactic phrases display recursivity, whereas this recursivity is assumed not to play a role in phonology. In music, however, recursive phrase structures are quite common, and this made us wonder why linguistic prosody would behave differently from both syntax and music. Moreover, recursion is found in all kinds of art and in nature as well. Section 4.2 gives several examples of different kinds of recursion; in the remainder of Chapter 4 we

---

<sup>1</sup> Sound examples from the experiments can be downloaded as mp3-files from <http://home.planet.nl/~schre537/sounds.htm> or [www.maartjeschreuder.nl](http://www.maartjeschreuder.nl).

search for evidence for the idea that phonology exhibits recursive structures as well. We conducted an experiment in which we studied an instance of phrasal structure. Thus we investigated whether or not edge-marking processes, such as early pitch accent placement, can be applied recursively to phonological phrases that are embedded in larger phonological phrases, and we show that recursion in phonological phrases should be admitted in the prosodic hierarchy.

The third experiment (Chapter 5) concerns the question whether differences in emotional speech are characterized by different modalities. In music the difference between sad and cheerful melodies is often indicated as a difference between a minor and a major key. We will indicate that we may also speak in a minor key when we are sad and in a major key when we are happy.

These prosodic subjects have in common that music theory can help out in the issues involved. All three subjects, concerning rhythm, phrasing structure, and intonation or melody, are basic parts of music theory as well. They are to be seen as building blocks of a bigger whole, in language as well as in music, building hierarchical structures out of sound. Without structure we cannot understand it. As for the simple example of the ticking clock, we hear structure in each part, and we connect it to the properties of the sounding signal, until we have reconstructed the entire piece of music or (spoken) text.





## Chapter 1

# Language and Music in Optimality Theory

### 1.1. Introduction<sup>2</sup>

Jackendoff and Lerdahl (1980) point out the resemblance between the ways both linguists and musicologists structure their research objects. This insight gave rise to the proposal of a formal generative theory of tonal music (Lerdahl and Jackendoff 1983), in which they describe musical intuition. Above all, insights from non-linear phonology (*cf.* Liberman 1975; Liberman and Prince 1977 among others) led to scores provided with tree structures, indicating heads and dependent constituents in the investigated domains. In this way, composer Lerdahl and linguist Jackendoff bring to life a synthesis of linguistic methodology and the insights of music theory. Gilbers (1987) shows that music theory in turn can be useful to describe linguistic rhythmic variability (*cf.* also Gilbers and Schreuder 2002). Further examples of musical and linguistic cross-pollination include among others Jacobson (1932), Guéron (1974), Liberman (1975), Attridge (1982), Oehrle (1989), Wallin (1991), Raffman (1994), Hayes and Kaun (1996), Hayes and MacEachern (1998), Patel (1998, 2003), Patel et al. (1996, 1997, 1998a,b), Repp (2000).

Liberman (1975) claims that in principle every form of temporally ordered behaviour is structured the same way (*cf.* also Gilbers 1992). If this claim is true, language and music should have much in common, since both disciplines are examples of temporally ordered behavior. In this chapter we offer additional arguments for this proposition. In both fields the research object is structured hierarchically and in each domain the important and less important constituents are defined. In Lerdahl and Jackendoff's music theory,

---

<sup>2</sup> This chapter is based on Gilbers and Schreuder (2002) which will also appear in two parts as Gilbers and Schreuder (in press) and Schreuder and Gilbers (in press). In Dutch it has appeared as Gilbers and Schreuder (2000).

these heads and dependents are defined by preference rules determining which outputs, i.e., the possible interpretations of a musical piece, are well-formed. Some outputs are more preferred than others. Preference rules, however, are not strict claims on outputs. It is possible for a preferred interpretation of a musical piece to violate a certain preference rule. This is only possible, however, if violation of that preference rule leads to the satisfaction of a more important preference rule.

This system of violable output-oriented preference rules in music theory has been very familiar to linguists since 1993, for a practically identical evaluation system, which uses similar well-formedness conditions, can be found in Prince and Smolensky's Optimality Theory (1993) (further OT). This theory, first introduced in phonology, owes a great deal to the work of Lerdahl and Jackendoff. Currently, it is a leading phonological theory and is expanding from phonology to other linguistic disciplines. In OT well-formedness conditions on outputs, constraints, also determine grammaticality. Here, too, the constraints are not strict, but soft, or violable. However, a crucial difference between Lerdahl and Jackendoff's violable constraints and OT's seems to be in the nature of the rule interactions. In Lerdahl and Jackendoff (1983), unlike standard OT, rules are not strictly ranked, because they apply with variable strength, and because sometimes several weaker rules can gang up on a stronger rule. The Lerdahl and Jackendoff theory is more like the theory of Harmonic Phonology, a predecessor of OT. Recent accounts of OT, however, have loosened the requirement of strict dominance. Through variants like constraint demotion (Tesar and Smolensky 1998) or the Gradual Learning Algorithm (Boersma and Hayes 2001), constraint rankings can vary to some extent (*cf.* Chapter 3). In this chapter we will show that in the present state of phonology the resemblances are even more striking than in the time of Lerdahl and Jackendoff (1983).

The remainder of this chapter is constructed as follows: section 1.2 of this introductory chapter is further devoted to the resemblances between Lerdahl and Jackendoff's music theory and OT, with subsections on structuring, conflicting preference rules, and boundary marking. Section 1.3 gives our conclusion in relation to the study of temporally ordered behaviour.

## 1.2. The resemblances between language and music

In their ‘Generative Theory of Tonal Music’ Lerdahl and Jackendoff (1983) describe how a listener (mostly unconsciously) constructs connections in the perceived sounds. The listener is capable of recognizing the construction of a piece of music by considering some notes/chords to be more prominent than others. This enables him for example to compare various improvisations on one theme and to relate them to the original theme, even if he does not know the original theme. It enables him to get to the bottom of the construction of a complete piece, as well as the constructions of the different parts of that piece. Where does a new part start? What is its relation to a preceding part? Which are the most prominent notes in a melody?

Our cognition thus works in a way comparable to how a reader divides a text (often unconsciously too) into different parts. A reader also distinguishes paragraphs, sentences and constituents. He structurally divides a text. What is the nucleus of a sentence? What is attributive and therefore less prominent?

The term ‘language’ as used in this dissertation has a very broad meaning. We mean any module of the language faculty which deals with hierarchical structure and which can be analyzed as consisting of deconstructable parts which stand in hierarchical relationships to each other, i.e. grammar. This contains at least syntax, morphology and phonology as it is represented in our unconscious knowledge.

In section 1.2.1 we will show what the resemblances are between language and music with regard to the division of the research object into smaller domains. Section 1.2.2 is about the resemblances in well-formedness rules, which are output-oriented, and which determine the main constituent and the dependent constituents for each domain.

### 1.2.1. Structuring

In music theory the musical stream of sounds is hierarchically divided into structural domains. Each domain contains some smaller domains, which in turn contain smaller domains. The smallest domain in music is the motif (built up out of notes), a short, rhythmic, melodic or harmonic building block, which is a recurrent element in the whole piece of music. It retains its identity when

elaborated on or transformed and combined with other material (Randel 1986: 513). Several motifs together form phrases, and phrases together may build up themes. A phrase, or period, is a kind of musical sentence, which concludes with a moment of relative tonal and/or rhythmic stability such as produced by a cadence, *cf.* section 1.2.2.4 (Randel 1986: 629). The realization of phrasing in performance is largely the function of the performer's articulation. A theme is a musical idea, usually a melody, that forms the basis for a composition of a major section of a composition. It can consist of a single phrase or several phrases together (Randel 1986: 844). It generally covers several measures and is regularly varied upon during the whole piece. In principle the listener is always able to recognize the theme, although it can be somewhat different each time. He reduces every occurrence of the theme to its underlying structure. The motifs and themes together determine the character of the piece of music. Several phrases or themes can form a section or verse, etc. By imposing this hierarchical structure on the entire piece, the listener is able to understand it. Figure 1 shows an example of the construction in the jazz original 'Tuxedo Junction'.

Figure 1 Tuxedo Junction

a. Motif



b. Theme or phrase



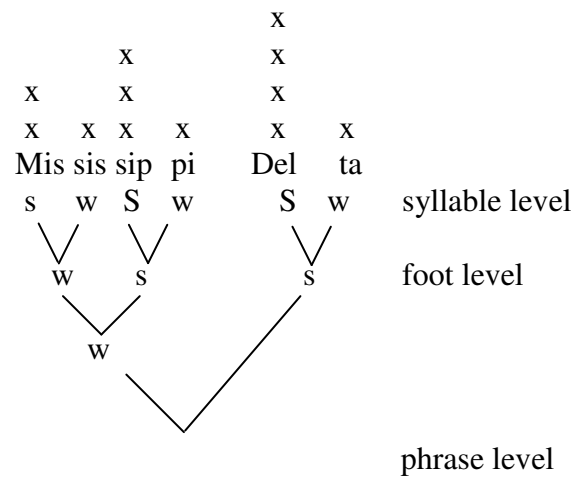
c. Section



Comparable domains can be found in language. The building block in language comparable to the motif in music is the morpheme (built up out of phonemes). Morphemes are joined together into larger meaning-bearing units: words, compounds, constituents (phrases),

etc. And just as we have a rhythmic division (*metrical structure*) in addition to a melodic division (*grouping structure*) in music, we can divide rhythm in language into syllables as well, united into feet, which are comparable to the musical measure. In language – as in music – this division of the sound signal into domains allows us to grasp the structure and to understand how to interpret the whole text. Figure 2 shows an example of a structured phrase in language. The height of the grids reflects the degree of stress and the tree diagram represents the relative strength between the syllables and feet.

Figure 2 Prosodic construction of a phrase (Prince 1983)



## 1.2.2. Conflicting preference rules

### 1.2.2.1. Evaluation of possible output candidates

In language (Prince and Smolensky 1993) as well as in music (Lerdahl and Jackendoff 1983) the head of each domain is chosen by means of well-formedness conditions. A coherent whole of such conditions (or constraints) indicates what is grammatical in language

and which mode of perception is optimal in music. In language for example one has to know which of two syllables in a foot is stressed and in music which chord of a certain sequence is the most prominent in the progression of the whole piece.

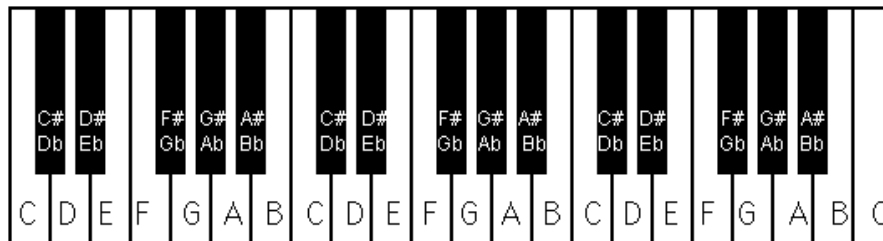
Possible candidates for every output form are evaluated by the constraints. These constraints can be contrasting and lay down opposite requirements on the output structure or interpretation to be preferred. Conflicts are thus solved by assuming differences in weight between the different constraints. In this way a weight hierarchy of constraints is arranged. One could compare this to traffic rules. Traffic coming from the right has priority, unless the traffic coming from the left is driving on a major road. This last rule, however, is overruled by the rule stating that one has to wait for a red traffic light. In traffic we are dealing with a collection of hierarchically ordered rules. Note that these rules are soft. They can only be violated in order to satisfy a higher preferred rule (minimal violability).

Linguistic constraints in OT are soft too. An output candidate can be grammatical, even if it violates constraints. As long as no better candidate comes up, the least bad candidate is the optimal one. Suppose we have a word with two syllables CVCVVC (*papaap*) and we have to determine on which syllable stress falls, given two relevant constraints: a positional constraint *i* (stress never falls on the last syllable) and one in which syllable weight plays a role, constraint *j* (stress falls on the heaviest syllable). The best output according to constraint *i* is then: *pápaap*, but *papáap* is the best according to constraint *j*. There is no output which satisfies both constraints. In a grammar conflicts like these are solved by a language-specific ranking of the constraints according to their importance. These universal constraints are not ranked in themselves, but in the grammar of a particular language they are strictly ordered. A language learner has to acquire the knowledge that in language X constraint *i* has priority over constraint *j*, while in language Y it can be the other way around.

The well-formedness rules in music theory are also potentially conflicting and soft. One of the conditions implies that a chord in a metrically strong position (for example the first beat in a measure) is more important than a chord that is not in such a position. A chord in a metrically strong position is preferred by the listener to act as most

prominent chord (the head) of the measure or the phrase, above all other chords in the same sequence. Another preference rule states that, given the tonality of the piece, all chords are harmonically unequal in their strength. In a piece in the key of C, the G-chord is harmonically more consonant than a B-chord. Thus there will be a conflict between preference rules if a B-chord is in the first position of a measure and a G-chord is in the last. Lerdahl and Jackendoff solve this kind of conflict by hierarchically ranking the preference rules. In our example the preference of a harmonically more consonant chord outweighs the preference of a metrically stronger chord, so that the listener will choose the G-chord as head and not the B-chord, given the key C.

Figure 3 Piano keyboard



An apparent difference between music and language is that Lerdahl and Jackendoff give only one ranking of well-formedness rules, while in OT a ranking of the universal constraints, in themselves unranked, has to be made for every language. Although Lerdahl and Jackendoff only offer one ranking for tonal music, one can imagine that, for example, prolongation of a melodic line is relatively more important in Eastern music than in Western music, while possibly in Western music relatively more weight is attributed to harmonic consonance of a piece. Perhaps differences in musical styles can be accounted for in the same way as for differences between languages (*cf.* also Patel and Daniele 2003, and Chapter 2).

In the next subsections we will discuss two examples of a conflict between positional and segmental markedness. In section 1.2.2.2 we present a linguistic example based on language acquisition data; in section 1.2.2.3 a comparable example in music is given.



### 1.2.2.2. A linguistic example of conflicting constraints: language acquisition

The language acquisition data in Table 1 prove that several kinds of markedness play a role in the acquisition of clusters. In this example we can see a conflict between segmental markedness and positional markedness in the realizations by the Dutch boy Steven of respectively *acht* ‘eight’ and *korst* ‘crust’.

Table 1 Cluster reduction Steven

age:	target word:	input:	realisation:
1;11	<i>acht</i>	/ɑxt/	[ɑt]
2;2	<i>korst</i>	/kɔrst/	[kɔs]

Data: Van der Linde (2001)

The dominating constraint in both cases is \*COMPLEX, a prohibition on consonant clusters in the output. Prince and Smolensky (1993) propose HMARG to indicate that in marginal syllable positions less sonorant segments are preferred to more sonorant ones. The child has arrived at a phase of its development in which the correspondence constraint MAX I-O, a constraint which demands that every segment of the input has a correspondent in the output, and therefore forbids deletion, is dominated by \*COMPLEX and HMARG. With the help of these constraints we get to the analysis in Table 2.

Table 2 Provisional OT analysis

constraints → /ɑxt/ candidates ↓	*COMPLEX	HMARG	MAX I-O
[ɑχt]	*!		
[ɑχ]		/χ/!	*
☞ [ɑt]		/t/	*

The constraint ranking in Table 2, however, wrongly predicts that the realisation of *korst* would be [kɔt]. We assume that Steven's realisation [kɔs] should be explained by the supposition that the difference between the syllable positions of /t/ and /s/ has its influence. HMARG is violated to satisfy a higher-ranked constraint with respect to positional markedness.

A straightforward CVC-syllable model and constraints like \*COMPLEX and \*CODA (syllables must end in a vowel) are not satisfactory for describing phonotactic restrictions and positional markedness relationships between segments in a Dutch syllable. We therefore copy a more complex syllable template in Figure 4 from Gilbers (1992). This model is based on a proposal in Cairns and Feinstein (1982), in which differences in positional markedness are stipulated, mixed with a proposal in Van Zonneveld (1988), in which an X-bar theory for syllable structure is developed<sup>3</sup>.

<sup>3</sup> Cairns and Feinstein indicate differences in markedness between consonant sequences like obstruent–liquid; obstruent–nasal. Unfortunately their model lacks sequences with fricatives such as in *schaap* [sχa:p] ‘sheep’.



Steven's realisations can be described by means of the tableaux in Table 4, based on the Constraint Demotion Algorithm for language acquisition by Tesar and Smolensky (1998). In Table 4a we see that before his second birthday Steven is in a phase in which segmental markedness (HMARG) dominates positional markedness (\*XSYLL, \*SAT), but that after his birthday positional markedness has become more important than segmental markedness. Finally, the correspondence constraints will dominate the markedness constraints. Phonological development is then completed.

Table 4 Analysis *acht* and *korst*a. table for *acht* (phase Steven (1;11))

constraints → /axt/ candidates ↓	*COMPL	HMARG	*XSYLL	*SAT	MAX I-O	*CODA
[axt]	*!	/χt/	*			*
[ax]		/χ/!			*	*
☞ [at]		/t/	*		*	

b. table for *korst* (phase Steven (2;1))

Constraints → /korst/ candidates ↓	*COMPL	*XSYLL	*SAT	HMARG	MAX I-O	*CODA
[korst]	*!	*		/rst/		*
[kør]			*!	/r/	**	
☞ [køs]				/s/	**	*
[køt]		*!		/t/	**	
[kørs]	*!		*	/rs/	*	*
[kørt]	*!	*	*	/rt/	*	
[køst]	*!	*		/st/	*	*

Notice that OT is not a theory on representations or models. Table 4 is based on the model in Figure 4, where /t/ is not a Coda because it is in an Appendix position, and /r/ is not a Coda because it is in the Satellite position.

In music conflicts also arise between positional and ‘segmental’ markedness. In the next subsection we give an OT analysis of a passage from Mozart.

### *1.2.2.3. A musical example of conflicting constraints: OT analysis of Mozart K. 331, I*

In music, similar to language, different preference rules can be arranged, in order to solve conflicts concerning the head of a domain. Segmental markedness has its musical equivalent in the hierarchical relationships between notes in a given tonality. Positional markedness is comparable to the strength differences between different positions in a measure.

With regard to segmental markedness, musical segments – like segments in language – keep hierarchical relationships with each other. The hierarchy of musical segments, the pitches, is connected to the tonality of the piece. In tonal music, every piece is based on a given scale (the key or tonality of the piece), which means that all notes are arranged around the most important notes in that scale; the melody usually ends in the tonic, the keynote of that scale.

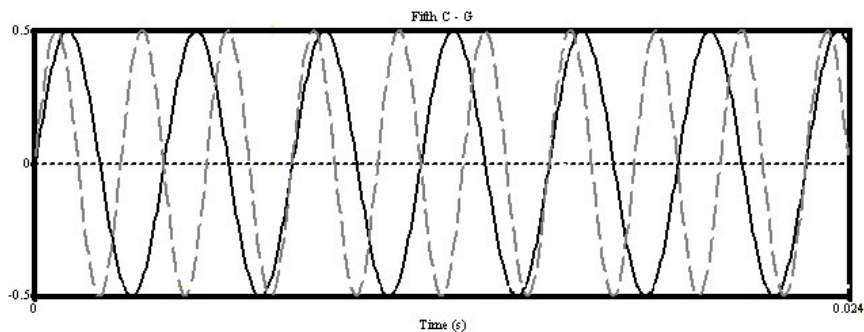
The tones of the scale can be combined in several ways, following each other in a melody, or harmonizing in chords. One harmony or succession sounds ‘better’ than the other. Intervals that are stable and do not require resolution are called ‘consonant’, more complex sounding intervals are called ‘dissonant’. Dissonant harmonies are regarded as having an instability that requires resolution to a consonance (Randel 1986: 197). Like sonority in language, consonance and dissonance are gradual concepts. The hierarchical division of pitches in a piece happens on the basis of the relative consonance (Lerdahl and Jackendoff 1977, 1983). A relative consonant tone in the key of the piece is higher in the hierarchy than a relatively dissonant tone.

That consonance and dissonance are not a matter of taste, but a matter of acoustics, is shown in Figure 5. Consonant intervals consist of a simple ratio, whereas dissonant intervals have a more complex ratio. The ratio in e.g. a fifth is 2:3, as illustrated by two wave forms

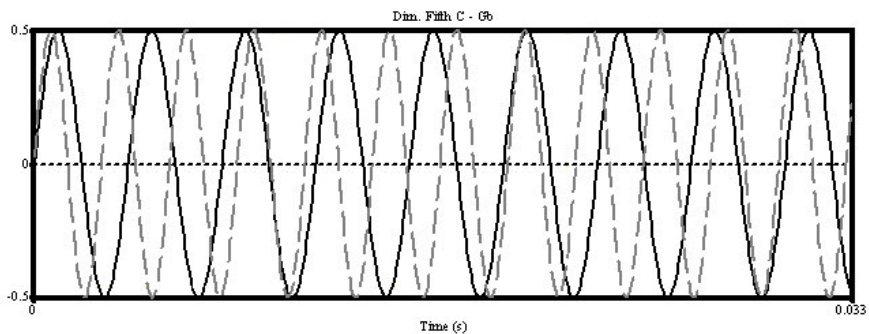
in Figure 5a. The three cycles of the G cross the C sinus in the zero boundary after two cycles of the C. The wave sinuses in Figure 5b, on the other hand, are much more complex; the sinus crossings do not intersect with the zero line anywhere. The extent of complexity of the wave ratios corresponds to the perception of relative consonance.

Figure 5 Consonance and dissonance

a. Consonant: perfect fifth C-G



b. Dissonant: diminished fifth C-Gb



In addition to segmental markedness, there is also positional markedness in music, just as we saw in Figure 4. The first position in a measure is stronger than the second, and in for example the 4/4-measure the third position is less strong than the first, but stronger than the second or the fourth.

Lerdahl and Jackendoff developed the so-called time-span reduction, a kind of tree and grid construction, based upon the

metrical structure and the grouping structure of a part of a musical composition, so as to reflect the hierarchical relationships between all pitches in relation to the tonality of the piece (see Figure 7). These relationships are determined by application of the preference rules, which determine the head in each domain. The head of a time-span *Z* is selected from the heads of the time-spans directly dominated by this time-span *Z*. The subordination relationship is transitive here: if *X* is an elaboration of *Y* and *Y* of *Z*, then *X* is also an elaboration of *Z*. Lerdahl and Jackendoff (1983) treat nine time-span reduction preference rules (TSRPR). Table 5 gives three examples of such rules.

Table 5 Time-span reduction preference rules

- TSRPR 1: Choose as the head of a time-span the chord (or the note) which is in a relatively strong metrical position (positional markedness).
- TSRPR 2: Choose as the head of a time-span the chord (or the note) which is relatively harmonically consonant (segmental markedness).
- TSRPR 7: Choose as the head of the time-span the chord (or the note) which emphasizes the end of a group as a cadence (comparable to the boundary marking effect of alignment constraints in language, *cf.* Table 7).

An example of a strong metrical position from TSRPR 1 is the first position in the measure. TSRPR 2 is connected to a hierarchy of chords based on harmonic stability. A triad tonic–tierce–fifth (c–e–g) is more stable than a seventh chord (c–e–g–b flat), while a seventh chord in its turn is more stable than for example a suspended fourth (sus4) (c–f–g). The optimal chord according to TSRPR 7 is the final chord, a chord which generally is built on the tonic, preceded by a dominant chord (see Figure 11a). In C the dominant is G. Each smaller group concludes with a chord suitable for a cadence. There are also ‘lighter’ cadences, however, indicating that a group is not definitely concluded, and that the melody will continue after the

cadence, moving to a next group. Often the sequence subdominant–tonic is used (the plagal cadence, F–C, *cf.* Figure 11c). The first three positions in the harmonic hierarchy are occupied by the tonic, the dominant, and the subdominant respectively.

As in OT the set of preference rules from music theory is hierarchical. TSRPR 2 is stronger than TSRPR 1; TSRPR 7 is stronger than TSRPR 1 and TSRPR 2 together. In Figure 6 we give the first movement from a sonata by Mozart.

Figure 6 Mozart: Sonata K. 331, I (Lerdahl and Jackendoff 1977)



For this part we can determine the heads by means of application of the TSRPR hierarchy. The first four measures from the piece form the first group. In measure 3 the  $A^6$ -chord (F#–E–A) is metrically the strongest chord, and thus the head. In measure 4 the E-chord (E–G#–B) is the head, because it marks the end of the whole first group of four measures. Now the head has to be chosen for the group which is formed by measures 3 and 4 together. Metrically speaking, the  $A^6$ -chord is still the strongest. But TSRPR 7 dominates TSRPR 1. In Table 6 we give an example of an OT-like musical analysis. Although the  $A^6$ -chord is metrically speaking in a stronger position than the E, the dominant TSRPR 7 prefers the dominant chord E as the cadence in this phrase.

Table 6 OT analysis

constraints → $A^6 - E$	TSRPR 7	TSRPR 2	TSRPR 1
Candidates ↓			
E			*
$A^6$	*!	*	



This choice has consequences for the tree in Figure 7, in which the E-chord dominates the A<sup>6</sup>-chord. The E-chord in its turn is dominated by the harmonically more consonant initial A-chord of the piece, and at the top of the hierarchy is the final chord of the whole group of eight measures, again an A-chord, because it is the head according to both TSRPR 1 and TSRPR 7.

Replacing all notes/chords which are chosen as heads of every time-span by gridmarks shows the resemblance to metrical phonological representations as proposed in Liberman (1975), Liberman and Prince (1977) and Hayes (1984) among others (see Figure 8). The underlined gridmarks (x) indicate *silent beats* (cf. Selkirk 1984). Silent beats are filled either by a rest or by lengthening of a preceding note. Note that metrical grids usually indicate stress differences, whereas this grid indicates prominence differences between chords, not stress. Obviously, the same kind of representations can be used to indicate differences in prominence.

The analysis shows that the beginning and end of the phrase are emphasized. TSRPR 7 dominates the constraints referring to segmental and positional markedness. In language, boundaries of a phrase may also be emphasized. In this way a stress shift as in *Mississippi Déltà*, realized in fast speech as *Mìssissippi Déltà* (Hayes 1984), can be described (cf. Visch 1989). We will examine this in the next subsection. For an elaborate experiment with regard to boundary alignment we refer to Chapter 4.

Figure 7 Time-span reduction (Lerdahl and Jackendoff 1977)

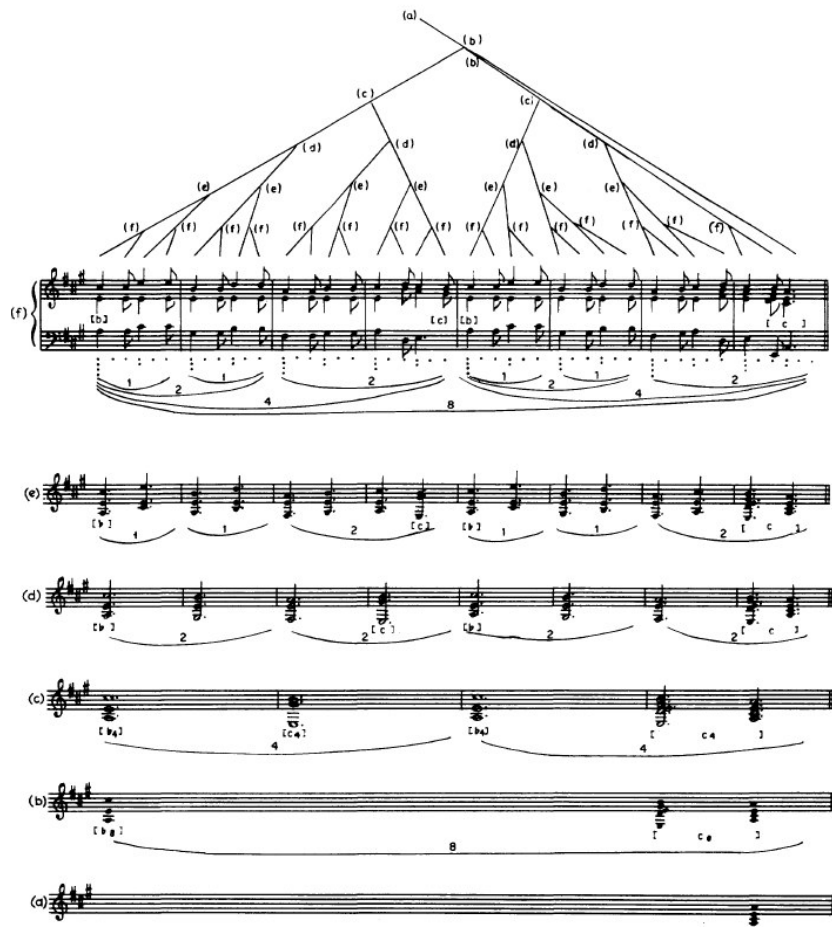
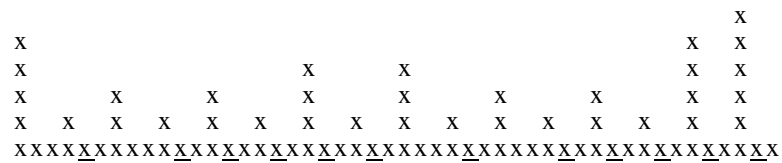


Figure 8 Grid of the time-span reduction in Figure 7  
(two adjacent phrases)



#### 1.2.2.4. Boundary marking

In both music and language, several processes can be considered to be boundary markers. Secondary stress shift (or ‘early accent placement’ as we call it in Chapter 4) and final lengthening are such processes. In OT so-called generalized alignment constraints are proposed for the analysis of boundary marking processes (McCarthy and Prince 1993b). All alignment constraints refer to constituent boundaries, and they have the following form:

Table 7 *Alignment*

Align (Cat 1, Edge 1, Cat 2, Edge 2) =  
 $\forall \text{ Cat 1 } \exists \text{ Cat 2 where Edge 1 of Cat 1 and Edge 2 of Cat 2 coincide}$

Alignment constraints prefer output candidates in which for example a constituent boundary coincides with a stressed syllable or in which a morphological boundary coincides with a phonological one.

A predecessor of alignment constraints for the controlling of rhythmical boundary marking in language is the Phrasal Rule of Hayes (1984). Hayes gives examples of preference rules for an ideal rhythmic structure in language: Eurhythmy rules. He attributes rhythmic shift to adjustments to ideal patterns for rhythmic sequences. The Phrasal Rule (PR) is one of these Eurhythmy rules. It implies that a grid is more eurhythmic if it contains two marks as far apart from each other as possible, at the second-highest level. The PR makes that the boundaries of the phrase are emphasized. Van Zonneveld (1983) called this phenomenon ‘Rhythmic Hammock’.

Table 8 Rhythmic Hammock in *individualistisch persoon*  
 ‘individualistic person’ (Visch 1989: 102)

constraints → <i>individualistisch persoon</i> candidates ↓	HAMMOCK	CORR
<i>individualistisch persóon</i>	*!	
☞ <i>individualistisch persóon</i>		*

In Table 8 the hammock pattern is visible the second candidate. This pattern is comparable to the grid pattern in Figure 8 for the music passage in Figure 7, where the extremes are marked by the highest grid columns. Because Hammock, like TSRPR 7 in music, is a dominant constraint, the second candidate in Table 8 wins. So like similarities in segmental and positional markedness we also see a great similarity between language and music in the way boundaries are marked. Hammock patterns are found in phonology as well.

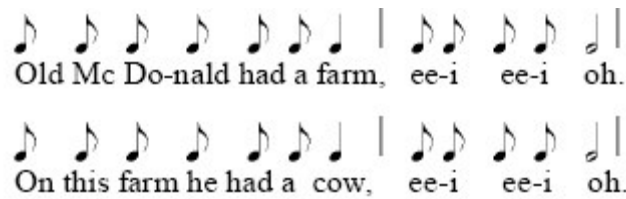
Another form of boundary marking which we find both in language and music is Final Lengthening (FL) (Lindblom 1978, Ladd 1996). FL is the phenomenon of lengthening of a note or a speech sound at the end of a phrase. According to experimental research by Lindblom (1978) in spoken Swedish the duration of the vowel [ɑ:] in [‘dɑ:g] is longer at the end of a phrase (Table 9a) than when the word is in another position (Table 9b).

Table 9 Final Lengthening in Swedish

- a. *finurlige Dag*      ‘ingenious Dag’  
 b. *Dag berättar*      ‘Dag tells a story’

In Table 9a the vowel is in final position and therefore it lasts  $\pm 55$  ms longer than in initial position, as in Table 9b. Figure 9 shows an example of FL in music, where it is a very common phenomenon.

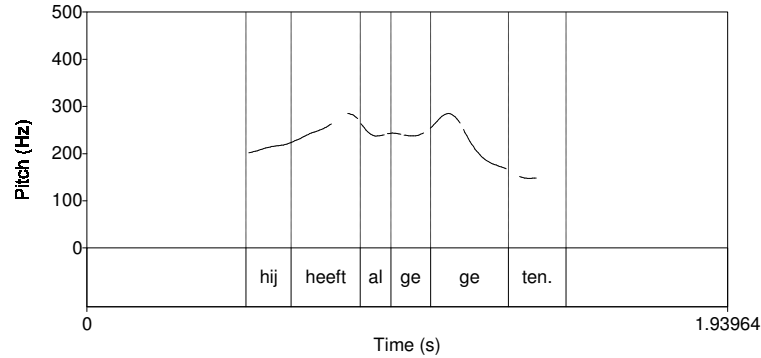
Figure 9 Final Lengthening in music (after Liberman 1975)



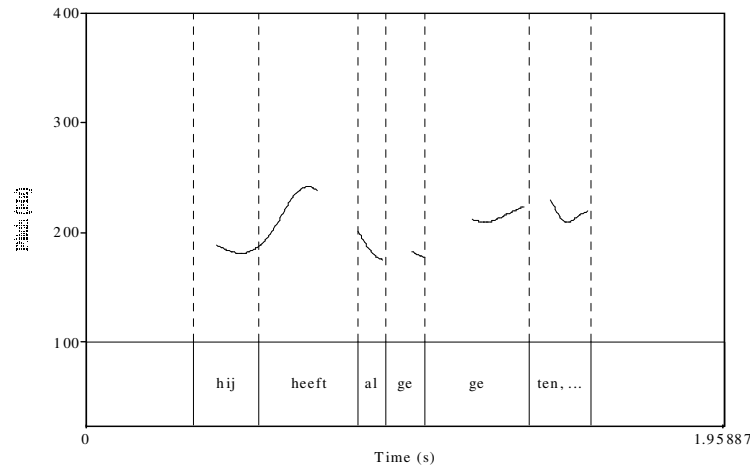
The last note of each phrase is lengthened indicating that the phrase is concluded. Another common phenomenon for marking boundaries in both language and music is deceleration of tempo (ritards) towards the end of phrases, as well as acceleration at the beginning of each melodic movement (Repp 1990, 1998). Patterns of tempo modulations often indicate a hierarchy of phrases, with the amount of slowing or phrase-final lengthening at a boundary reflecting the depth of embedding (Todd 1985, 1989, Palmer 1989, 1997, Repp 1990, 1998, 2002, H. den Ouden 2004). One can see that this gradation in FL occurs in Figure 9, as the note 'before the comma' is lengthened compared to the preceding notes, but less than the final note of the phrase.

In addition to rhythmic phenomena, intonation patterns are used to mark boundaries. In language, intonation marks groups such as syntactic constituents and phonological phrases. In a similar way intonation marks, for example, the differences indicated in writing by full stops and commas. A full stop in a declarative sentence is often the equivalent of a strong pitch fall in prosody, while a comma is comparable to the intonation pattern in which the tone is suspended somewhere 'in between', to indicate that the sentence is to be continued (Swerts 1994, Van Donzel 1997, 1999). The contours in Figure 10 reproduce this difference.

Figure 10 Intonation patterns (Schreuder 1999)



a. *Hij heeft al gegeten.* 'He has already eaten.'



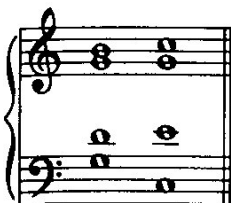
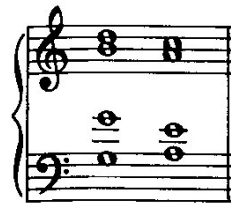

b. *Hij heeft al gegeten, (maar hij wil toch nog een koekje.)*  
'He has already eaten, (but he still wants another cookie.)'

In Figure 10a the intonation contour moves downward towards the end, creating a 'final fall', and in Figure 10b, the 'comma intonation', the tone is suspended in between (it is rather high in this example), the 'continuation rise'. Thus in Figure 10b the sentence cannot be complete, something has to follow this boundary. The boundary tone of a question often rises to the top of the speaker's

range, although in the case of question intonation non-finality may not be the reason for the rise of tone (Gussenhoven 2002, 2004).

Intonation in language equals phrasing in music.<sup>4</sup> It causes the music to ‘tell a story’, similar to the way intonation does in language. Phrases are formed in which tension is built up or reduced, ending in a cadence, a melodic or harmonic configuration that creates a sense of repose or resolution (Randel 1986: 120). This is properly comparable to ‘comma intonation’ and ‘full stop intonation’ in language: the comma indicates prolongation, the full stop completion.<sup>5</sup> A full stop is comparable to the ‘full cadence’ (the end of a phrase or piece) in music, i.e. the sequence of G-C in the key of C, as in Figure 11a. Phrases and pieces prefer ending in the tonic, here C, mostly low.

Figure 11 Cadences

a. Full cadence	b. Deceptive cadence	c. Plagal cadence
		

A comma is comparable to a chord change in which the phrase does not end in the tonic, but e.g. in the fourth step of the scale (a ‘deceptive cadence’, as in Figure 11b), F in the key of C, so higher. It therefore does not sound completed, and another phrase, resolving in the tonic, will ideally follow. While in speech intonation the non-final boundary tone is mostly higher than the final boundary tone, in music this is no more than a tendency, because both boundary tones

<sup>4</sup> The term ‘intonation’ in music is reserved for tuning. We do not use the meaning ‘tuning’ in this dissertation.

<sup>5</sup> Lerdahl and Jackendoff (1983) describe the difference between intonation patterns expressing prolongation and intonation patterns expressing completion. Prolongation is worked out in the prolongation reduction of the pitch structure.

can be the same tone. Depending on the melodic lines, and the harmonic progression of different cadences, the suggestion of a comma can nonetheless be evoked, as exemplified in e.g. the Marseillaise.

In Figure 12 we show a musical example of phrasing: ‘question and answer’, or ‘antecedent and consequent’. One phrase (the answer or consequent) follows the other phrase (the question or antecedent) and is also a reaction to it. The two phrases often have the same or similar rhythms, but have complementary pitch contours, e.g., a rising contour in the first and a falling contour in the second. An example is given in the first three measures of Mozart’s 40<sup>th</sup> Symphony in G Minor, K. 550:

Figure 12 Mozart K.550 (fragment): antecedent and consequent



This ‘question-answer intonation’ has a way of indicating grouping boundaries that is parallel to the patterns of full stops and commas. Again it is very similar to the patterns appearing in language. Questions have the tendency to end ‘upward’, while answers, comparable to sentences with full stop, tend to show a strong final fall. In fact, the example in Figure 12 shows this relationship at two levels simultaneously. This antecedent-consequent pair is followed by a similar pair one tone lower. At the same time, the two pairs are also related to one another as antecedent and consequent (Randel 1986: 42).

In this section, we showed that language and music have many similarities both on a representational level and in the sphere of preference rules. It seems that output-oriented preference rules do not specifically hold for only one discipline. In Chapters 3, 4 and 5, we will see that insights from music theory can be very useful in phonological issues.



With this in mind, we want to add a remark about the division into temporally organized elements, such as segments, accents, rhythm, or chords, and holistic patterns like intonation contours of phrases and melodies, all of which has to do with cerebral hemisphere specialization. In general, language processing is known to be situated in the left hemisphere. Only intonation is one of the few properties of language that are processed in the right hemisphere. Platel et al. (1997) and Stowe et al. (2005) point out that music perception is also located in both hemispheres; temporal patterns, like rhythm and chords, are located in the left hemisphere, and sequences of tones, i.e. melodies, and timbre, in the right. So intonation is literally the melody of language. The left hemisphere seems to be specialized in linear processes and consequently in analyzing temporal structures, of which rhythm, accents, segmental and positional hierarchies, etc. are examples. It was found that rhythm in music is processed by Broca's area, one of the neurological areas which are specialized in language processing. The right hemisphere, on the other hand, analyses in a holistic manner. It processes complex relationships and perceives patterns as units, instead of as sums of individual parts. These findings are highly controversial, however.

In spite of the controversy on this subject, it should be noticed that these alleged differences in processing of the two hemispheres reflect the differences between Lerdahl and Jackendoff's time-span reduction and the prolongational reduction. The prolongational reduction also analyses parts of a melody as larger units, not in a bottom-up fashion like the time-span reduction, but top-down. This might be new evidence for Lerdahl and Jackendoff's separation of the time-span reduction and the prolongational reduction (not elaborated in this dissertation. This separation has psychological relevance. It also shows that their theory, and especially OT, may give a good model for the way our brains work. We should keep in mind that OT has its source in connectionism. Connectionist models were developed in an attempt to construct a model that closely resembles the structure of the human brain (*cf.* Gilbers and De Hoop 1998).

### 1.3. Summary and Conclusion

In this chapter we followed Lerdahl and Jackendoff (1983), who observed a resemblance in the way musicologists and linguists structure their research objects. This observation led to their book ‘A Generative Theory of Tonal Music’, in which they describe music by means of a linguistic methodology: they used trees and grids, very common tools in syntax and phonology. Using these representations, they could visualize the hierarchical structures of which music is built up.

The methodology of preference rules Lerdahl and Jackendoff introduced to describe the way of achieving the ideal interpretation of a musical piece was followed ten years later by Prince and Smolensky (1993) in their Optimality Theory. The violable OT constraints show a striking similarity to the preference rules for music in defining the optimal output or musical interpretation. In this chapter we pointed to this similarity and we compared some of the musical preference rules to OT constraints.

We showed in this chapter that language and music also have much in common with respect to psychological assumptions and structural properties. In both disciplines the ‘grammar’ imposes hierarchical structures on the sound signal. In both language and music, preference rules for ideal outputs indicate the head constituent and the dependent constituents of every part of the hierarchical structure. Together the preference rules or constraints indicate what is grammatical in language and which way of listening is optimal in music. Moreover, in both theories the preference rules are soft and potentially conflicting, which gives the theories their descriptive power.

We gave a musical and a linguistic example of a conflict between positional markedness constraints and segmental markedness constraints. In music this conflict must be solved in order to decide which chord is the most prominent and will survive in the reduction. In language the outcome of the conflict is crucial to which segment is most prominent and will survive the simple system of a child acquiring language.

The domains in musical and linguistic structures are analogous. Both are deconstructible into smaller building units. The boundaries of these domains are the areas of several processes, many of which

are involved in boundary marking mechanisms. Common phenomena of boundary marking in both language and music, such as phrasal accents that shift to the left boundary of a phonological phrase, and also final lengthening at the right boundary, are regulated by alignment constraints. In language as well as in music, the initial element is important, at all levels. Final elements are important as well, which is illustrated by resemblance of cadences in music to the final fall or rise in linguistic prosody. This should be seen as the result of satisfaction of so-called generalized alignment constraints. With respect to rhythm, similar restructuring processes occur, which seem to be the result of constraints referring to the Obligatory Contour Principle (OCP), a prohibition on adjacency of identical elements (McCarthy 1986). These constraints take a prominent position in the constraint ranking.

In our view, the observation that language and music show so many similarities strengthens the hypothesis that the same structures and principles hold for all temporally ordered behavior (*cf.* Liberman 1975; Gilbers 1992). In addition we can refer to research by Lasher (1978), who describes patterns in ballet in a similar way to our description of language and music in this chapter. In her research of dancing patterns the main movements are also distinguished from dependent movements, for every part of the hierarchically structured research object. It is the way in which our brain works: our cognitive system structures the world surrounding us in a particular way in order to understand everything in the best way.

On the basis of these resemblances we will show that insights of music theory can help out in phonological issues. Three of such issues are subjects of the experiments in this dissertation: variable rhythm, variable phrasing structure, and emotional intonation. In the remaining chapters of this dissertation we report on these experiments.



## **Chapter 2**

# **Rhythm**

### **2.1. Introduction**

In the foregoing chapter we have seen that language and music have similar kinds of structures and processes, both at the level of phrasing and at the rhythmic level. The remainder of this dissertation is about phrasing and rhythm phenomena in speech, while we keep musical phrasing and rhythm as the background motivation for several speech phenomena. The current chapter gives a (non-exhaustive) overview of the literature on rhythm, and provides the background information for both Chapter 3 on rhythm in fast speech, and Chapter 4 on recursive prosodic phrasing. We do not intend to contribute to the theoretical discussions involved in the rhythmic issues discussed in this chapter. The issues discussed are the following: in the first section we will describe what rhythm is. In the second section we will make the distinction between rhythm and meter. Section 2.3 shows that languages differ as to how their rhythm units are classified. Section 2.4 describes some variable rhythm patterns in speech, and the rules that seem to play a role in this variability. In section 2.5 we will point out that time is a major factor in rhythmic patterns, which also will turn out to be the core explanation for our observations in Chapter 3. Firstly, we will answer the question “what is rhythm?”.

Rhythm is everywhere, in the world within us and in the world around us. Rhythm is in our heartbeat, our breathing, and our stride, but also in the tides of the sea, the seasons and the movements of the earth itself. These are all movements in a rhythmic fashion. It turns out to be really hard to do things non-rhythmically: when people are asked to tap their fingers on the table irregularly, some recurrent pattern will appear (Fraisse 1982). This all suggests that rhythm is at the heart of nature, at least in natural events in which time,

movement, or visual patterns are involved. Language and music are just two of these rhythmical behaviors.

Rhythm usually implies some kind of succession of strong and weak elements of the same type, where the events occur with some repetitive structure, perceived at a constant rate, as a recurrent pulsation. Because the perception of a rhythm pattern needs some frame of reference to which the rhythmic parser can be set, it must be a series of at least three events or two (time) intervals to form a rhythmic pattern (Couper-Kuhlen 1993, Auer, Couper-Kuhlen and Müller 1999). A sequence of two events implies a single interval, which would not give the listener any reference to the internal relationships within the rhythmic pattern and the level on which regularity could be perceived. A sequence of three events gives us a succession of two (time) intervals, and this would give the listener an indication of the tempo and the regularity to be perceived.

Rhythm can be a pattern of any sequence, but mostly time and accentuation are involved (Randel 1986: 700-705). It can be thought of as the division of a temporal flow into perceptual groups. The grouping is what makes it rhythmical; without grouping the elements, it would be just a continuous flow of sound, going by unnoticed after a while. The elements are grouped by emphasizing some elements and by boundary marking. When we listen to sequences of pulses, any element that is louder, longer in duration, higher in pitch, or otherwise different from the rest, is perceived as leading the group of pulses. Combinations of differences in loudness, pitch and duration can lead to complex rhythmic groupings. The end of a rhythmic group is usually marked by a boundary marker, such as lengthening of the final element, deceleration of the tempo, or a pause after the final element.

Rhythm is only heard if a sequence of pulses is neither too rapid nor too slow (Woodrow 1951, Fraisse 1963, Randel 1986). The minimum time between pulses of which successiveness and order are perceivable is about 0.1 second, whereas the maximum beyond which groupings do not form is about 3 seconds (Allen 1975). Phonological rhythms generally work at close range, within the phrase and not beyond the sentence (Allen 1975, Couper-Kuhlen 1993, Auer, Couper-Kuhlen and Müller 1999).

Perceiving rhythm is imposing some rhythmic structure on a sequence. Even where it does not exist in fact, rhythm is heard. A

nice example is the ticking of a clock, as already mentioned in the Introduction of this dissertation (Bolton 1894, Fraisse 1963). The ticking is imitated by people by assigning the ticks different sounds, as if the two sounds differ. This indicates that we perceive the sounds differently, the tick sound emphasized with the high pitch of the intrinsically higher vowel /i/, the tock sound with a lower and therefore less prominent vowel /ɔ/. And by differentiating between the sounds, we automatically group them into groups of two, sometimes with a pause between the groups. This grouping behavior influences our perception of many temporally ordered entities.

Linguistic rhythm does not carry much information of its own, other than helping to guide the hearer's attention. Speech rhythm functions mainly to organize the information-bearing elements of the utterance into a coherent package, in order to make the informative elements temporally predictable in speech communication (*cf.* Chapter 3). Without rhythmic organization, the linguistic message would be difficult to transfer (Allen 1975).

## 2.2. Rhythm and meter

Rhythmic groups in speech and music can be regular or irregular patterns of different or similar elements. They consist of temporal relationships in a sequence of long and short sounds and silences, a free and creative ordering of time values (Broeckx 1967, Randel 1986).

Perfectly regular, recurrent rhythm constitutes meter (Randel 1986: 489, 702). Meter is the rhythmic level at which the conductor of an orchestra waves his baton and at which people dance. The term beat is sometimes used in a more abstract way for an arbitrary level of the metrical hierarchy. In music the term 'tactus' is used to indicate on which specific level the music is to be counted. It refers to the most salient periodicity or metrical level. This occurs between about 40 and 300 beats per minute, with a preference for a tempo of around 100 beats per minute, the so-called 'preferred rate' – a time interval of 600 ms (Fraisse 1982). In the case of speech, this level is also called the level of scansion. Note that for popular music, music software such as Steinberg Music and Adobe Audition take 120 bpm as the default setting, and in techno music the default rate is even 140

bpm or higher (Noys 1995, Reynolds 1999). We can only speculate that the preferred rate, parallel to the tempo of society, is changing to a faster beat. On the other hand, this tempo difference could be a distinction between popular and classical music.<sup>6</sup>

Meter thus involves isochronic time intervals between rhythmically prominent elements. The question now arises how regular rhythm must be in order to be perceived as metric. Experimental literature about isochrony in music and speech fails to confirm its existence; intervals appear to be highly variable (cf. e.g. Bolinger 1981, Dasher and Bolinger 1982, Dauer 1983, Couper-Kuhlen 1993, Laver 1994, Terken and Hermes 2000, Fox 2000 among others for speech, e.g. Honing 2002, and Handel 1993 for music). Isochrony seems to be subjective rather than objective: listeners tend to expect intervals to be isochronic, and they tend to adapt their perception to their expectations. In order to perceive regular patterns, absolute precision is not required and considerable latitude is allowed without destroying the sense of isochrony. A tendency exists to underestimate the duration of long time intervals, whereas the duration of short ones may be overestimated (Allen 1975), with the result that shorter and longer intervals may be perceived as having equal durations (see also our results of Chapter 3).

Whereas rhythm is built on physical stimuli in the acoustical surface, such as differences in loudness, duration, or contour, meter is an idealization, a psychological percept of the rhythmic stimuli. Metric accent in music stems from the listener's active engagement with the music: his mind must perform a number of interpretive tasks to hear it (Couper-Kuhlen 1993, Auer, Couper-Kuhlen and Müller 1999). This can be shown if we vary the context in which a rhythmical sequence is perceived. In Figure 13 we see two melodies, where the only rhythmical difference between Figure 13a and Figure 13b lies in the first two measures. In the first two measures of Figure 13a the notes have a length of three counts, dividing the measure in two, which gives the impression of a bipartite meter with groups of three quavers, as in a 6/8 measure. Figure 13b on the other hand starts with three notes of two counts each measure, which causes the

---

<sup>6</sup> Interestingly, hip hop music, which is a kind of (fast) speech, has a tempo in the same range as we find for speech, as described in Chapter 3: 80 to 120 bpm.



following quavers to be perceived as grouped in two by two, as in a 3/4 measure (London 2001).

Figure 13 Metric grouping



Meter is strongly predictive: metric patterns usually remain constant and are thus a reliable basis for anticipation when subsequent events will occur, while rhythm is not a reliable basis for anticipating subsequent events, although durational patterns may be repeated (Randel 1986: 489, Sachs 1953). Meter is also continuous, which means there are no gaps between successive groups. In the case of rhythm there are often gaps between successive groups.

The musical literature makes a clear distinction between rhythm and meter (Randel 1986: 702). In the linguistic literature, on the contrary, confusion of terminology arises, or one could just say that the terms are used differently. Mostly when phonologists speak of rhythm, they actually mean meter. Meter is seen as a subset of rhythm, namely as regular rhythm, which may in some cases be correct. The source of the confusion may lie in the fact that the perception of meter is often based on rhythmic events, and language, except for metric poetry, is usually not strictly regular. The distinction between rhythm and meter in language is therefore less easily made. The result of this terminological confusion can also be seen in Chapter 3 of this dissertation, where we first tried to find evidence for rhythmical shifts, but later found out that the perceived rhythms were only based on timing intervals, i.e., induced meter from signals differing in tempo. Although rhythm and meter are evidently distinct from each other, we will treat meter as the abstraction of regular rhythm in this dissertation, to fit in with the literature on linguistic rhythm, and also because the term ‘meter’ is already reserved for ‘meter’ as in prosodic foot structure (Kiparsky 1975), which differs subtly from the musical ‘meter’. The reader should keep in mind the distinction we made above.

The perception of rhythm and meter in speech also depends on the structural levels of rhythmic organization. Languages differ in the levels on which rhythm is based. In the next section we will therefore introduce the typology of rhythm classes in languages.

## **2.3. Rhythm units**

### **2.3.1. Rhythm typology**

In linguistic rhythm a central role is played by the syllable as a structural unit. The syllable can be interpreted as a unit of length in its own right, apart from the length of its segments. The length of the segments are in fact determined by the syllable structure and syllable weight, as can be seen in e.g. lengthening in open syllables in Dutch. Syllables are combined into feet ( $\Sigma$ ). The foot is fundamental in determining the positions of stressed vs. stressless syllables within words and larger strings. The structure of a foot can be characterized as consisting of a string of one relatively strong and any number of relatively weak syllables dominated by a single node (Nespor and Vogel 1986).

The different units formed by the prosodic hierarchy (Nespor and Vogel 1986) influence the rhythm patterns that are allowed. All levels of the prosodic hierarchy require their own rules of stress and accent assignment. Moreover, the boundaries of each prosodic level cause pre-boundary lengthening and pauses of relative strengths.

Traditionally, languages are classified according to three different rhythm classes: stress-timed languages, syllable-timed languages (Pike 1945), and mora-timed languages (Trubetskoy 1939: 171). The classification of languages into the different classes has long been controversial. More recently, however, more and more evidence has been found for the correctness of most of the classifications, on the basis of acoustic, perceptual and psycholinguistic investigations (among others Cutler et al. 1997, Ramus et al. 1999, 2003), although it is still not clear whether this classification into three groups is exhaustive, or that it should contain more categories, or that it rather should be seen as a continuous scale. Dutch and English are classified as stress-timed languages, French and Spanish are

exemplary for syllable-timing, and Japanese is the best-known example of a mora-timed language.

The acoustic properties of the different rhythm classes have been thoroughly investigated. Nevertheless, until recently most research has failed to confirm the existence of different types of isochronous intervals in spoken language (among others Bolinger 1965, Abercrombie 1967, Roach 1982, Dauer 1983). The impression is that in stress-timed languages two stressed syllables occur at approximately equal intervals of time, which implies isochrony between stressed syllables and rhythms of alternation. This would mean that the intervals between the stressed syllables in the title of this dissertation *pro'sodic 'processes in 'language and 'music* are equal. Stretches of unstressed words or syllables are therefore compressed, while adjacent stressed syllables are rhythmically separated by 'silent beats'.

In syllable-timed languages, on the other hand, one gets the impression that each syllable tends to be given the same space and that rhythm emerges because syllables (rather than stresses) occur at equal intervals of time. All syllables in a single phrase would take approximately the same time and make up rhythms of succession. Thus, in for instance *c'est absolument ridicule* 'it is absolutely ridiculous' all syllables would sound as isochronous.

In mora-timing the syllable would not be the basic rhythm unit, but the mora, the weight-bearing part of the syllable, is the unit of consistent length, and thus the rhythmical basis of the utterance. Morae consist of a short vowel and the preceding consonant. Some consonants, mostly nasals, can also serve as syllable nuclei on their own, representing one mora. Thus, the Japanese word *sensei* 'teacher' is quadrimoraic (*se-n-se-i*). In mora-timed languages the succession of elements of the same length is said to make up the feeling of a staccato rhythm (Fox 2000).

Dauer (1983) observed that for the division into stress-timed rhythm and syllable-timed rhythm some distinctive phonological properties play a role, mainly syllable structure and vowel reduction. Stress-timed languages have a greater variety of syllable types than syllable-timed languages and syllable weight plays a major role for stress assignment in these languages. With respect to vowel reduction, unstressed syllables in stress-timed languages usually have reduced vowels, are shorter or can in fact be absent. Recently, Ramus

et al. (1999) found that, besides these phonological properties, differences in vowel/consonant segmentation characterize the rhythm classes. These and other phonological and phonetic features combined with one another give the impression that some syllables are far more salient than others in stress-timed languages, and that all syllables tend to be equally salient in syllable-timed languages.

### 2.3.2. Musical rhythm parallels the linguistic rhythm typology

A number of musicologists and linguists have claimed that the prosody of a composer's native language can influence the structure of his or her instrumental music (e.g. Abraham 1974, Wenk 1987). However, this was never satisfactorily supported by experimental evidence. Patel and Daniele (2003) found a way to compare rhythmic patterns in the English and French languages and in classical music. They claim that spoken prosody leaves an imprint on the music of a culture.

Patel and Daniele (2003) used the speech materials and results from a linguistic study on rhythm structure by Ramus (2002). This study showed that British English, as a stress-timed language, had significantly higher values for variability of vocalic durations than French. The musical materials consisted of instrumental music of a relatively recent musical era, of composers who were native speakers of British English or French, who lived in England or France. The measurements were made on scores, not on performed music.

Their results show that there is much overlap between the English and French composers, but on average a robust difference emerges, which is in the same direction as the linguistic difference. The difference for music is smaller than that for speech, however, which is said to reflect within-culture variability. Nonetheless, they proved that the musical rhythmic differences of certain cultures parallel the rhythmic differences between native languages of those cultures.

In stress-timed languages such as Dutch and English, some rhythmic processes have been observed since decades. Rules like the 'rhythm rule' are common knowledge in the phonological literature on stress and accent (among others Liberman and Prince 1977, Hayes 1984, Selkirk 1984 among others). Before we turn to the empirical findings on rhythmic phenomena of this kind, the next section will first introduce the theories on the mechanisms which bring about

these phenomena. We will add some new data to these theories to show that more mechanisms seem to play a role and we will try to put these phenomena into explanations within the framework of Optimality Theory.

## 2.4. Variable speech rhythm

### 2.4.1. Eurhythmmy in speech

Rhythmic patterns, in speech as well as in music, can sometimes change due to several factors. In Chapter 1 we saw the effect of an asymmetrical principle of Eurhythmmy in language, the Phraseal Rule (PR). Hayes (1984) formulated two more, symmetrical, principles of eurhythmmy. Because Hayes' Eurhythmmy rules are set up for outputs, his theory can be seen as a kind of predecessor of OT. They can thus easily be used as OT constraints. The first of the Eurhythmmy rules is the Quadrisyllabic Rule (QR), which demands that a secondary accent in a phrase is ideally at a distance of four syllables from the main accent. For longer phrases it is the case that at the so-called 'level of scansion', the level immediately under the level on which the gridmark of the main accent is situated, the major beats are at a distance of two syllables from each other. In the traditionally cyclically derived structure of the phrase *Mississippi Déltà* this is not the case. The secondary accent on *sip* is just two syllables apart from the main accent on *Del* here. The QR thus prefers a secondary accent on the first syllable of the phrase and indeed the phrase is often pronounced as *Mississippi Déltà*. In this instance a conflict arises between the QR and a correspondence rule which, on the basis of *Mississíppi*, prefers the accent to fall on the penultimate syllable. We will start from the assumption that different rhythmic structures imply different constraint rankings, following Schreuder and Gilbers (2004b). In fast speech the QR is dominant. Note, however, that we will arrive at an alternative approach towards the end of this chapter, which will be elaborated in Chapters 3 and 4.

Table 10 OT analysis

constraints → <i>Mississippi Delta</i> candidates ↓	QR	CORR
☞ <i>Mississippi Déltà</i>		*
<i>Mississìppi Déltà</i>	*!	

The QR divides language into a kind of 4/4 measure. In tonal music this is also a common kind of measure, in which – as was said in the foregoing section – the first count in the measure is important and the third count receives a lighter accent. This last effect is seen again in the Eurhythm Rule known as the Disyllabic Rule: immediately under the level on which the major beats are at a distance of four syllables from each other, the major beats are ideally two syllables apart. A rhythmic pattern satisfying these Eurhythm rules shows a regular alternation of strong and weak elements at all levels (*cf.* TSRPR 1, Chapter 1). In Table 8 we see the interaction of the different Eurhythm Rules with the OT constraint CORRESPONDENCE.

Table 11 Eurhythmy Rules in *twenty-seven Mississippi legislators*

constraints → <i>twenty-seven Mississippi legislators</i> candidates ↓	PR	QR	DR	CORR
<pre>               X             X   X   X   X   X   X           X   X   X   X   X   X         X X X X X X X X X X X X <i>twenty-seven Mississippi legislators</i> </pre>	*!	*		
<pre>               X             X   X   X   X   X   X           X   X   X   X   X   X         X X X X X X X X X X X X <i>twenty-seven Mississippi legislators</i> </pre>	*!			**
<pre>               X             X   X   X   X   X   X           X   X   X   X   X   X         X X X X X X X X X X X X <i>twenty-seven Mississippi legislators</i> </pre>		*!*		*
<pre>               X             X   X   X   X   X   X           X   X   X   X   X   X         X X X X X X X X X X X X <i>☞ twenty-seven Mississippi legislators</i> </pre>				**

In the first output candidate both the PR and the QR are violated, and it is therefore rejected as an optimal candidate. The second candidate is also rejected on the basis of the PR. It violates CORR twice, but it was out of competition already because of the fatal violation of the PR. The third candidate is fine with regard to the PR, yet it violates the QR twice: once between the main stress on *le* and the syllable *sip*, with a distance of only two syllables, and once between *sip* and *twen*, which has an overlong interval of six syllables. The first violation of the QR is already fatal, because a better candidate, which satisfies the QR, can be found. Candidate 4 is the optimal candidate of the candidate set in this constraint ranking, in spite of its two violations of CORR. All three – higher-ranked – eurhythmy rules are satisfied by this output candidate.

These rhythmic preferences in speech show that the rhythmical structure may change. The question is what kind of properties influence these changes. In Chapter 3 we try to answer the question whether speech rate can lead to rhythmic restructuring. In the next section we will first look at some other varieties of variable linguistic rhythm.

#### 2.4.2. Triplet rhythm in trochaic Dutch

Besides the eurhythmy rules, a very well-known phenomenon in speech is clash avoidance (Prince 1983): clashes of strong, stressed, or accented syllables are avoided to prevent the rhythm from becoming ‘staccato’. This phenomenon is formulated in the constraint \*CLASH. A preference seems to exist for beats that are more evenly distributed over the phrase. In Table 12 we see that the distribution of strong syllables over the word *bijstanduitkeringsgerechtigde* ‘person entitled to social security’ in an andante tempo is different from the distribution in an allegro tempo. The phrase gets a triplet-like rhythm at allegro tempo. Gilbers (1987) notices that in fast speech the fourth syllable is stressed more than the third. In andante speech, however, the first syllable in *uitkering* gets more stress than the second.

Table 12 Rhythmic structure *bijstanduitkeringsgerechtigde* (Gilbers 1987)

a. andante	<i>bij</i>	<i>stand</i>	<i>uit</i>	<i>ke</i>	<i>rings</i>	<i>ge</i>	<i>rech</i>	<i>tig</i>	<i>de</i>
	s	w	s	s	w	w	S	w	w
b. allegro	<i>bij</i>	<i>stand</i>	<i>uit</i>	<i>ke</i>	<i>rings</i>	<i>ge</i>	<i>rech</i>	<i>tig</i>	<i>de</i>
	s	w	w	s	w	w	S	w	w

Neijt and Zonneveld (1982) and Van Zonneveld (1983) have argued that Dutch is a trochaic language. In OT terms this means that the constraints RHYTHMTYPE=TROCHAIC (RT=T), a foot consists of a strong syllable followed by a weak one, and FOOTBINARITY (FTBIN), a foot consists of two syllables (or two morae), are high in the



constraint ranking for Dutch. The question now is how these constraints account for the triplet rhythm in Table 12b.

Gilbers and Jansen (1996) formulate an elaborate OT grammar for Dutch stress patterns, partly based on Nouveau (1994), Van Oostendorp (1995) and Kager (1994). The relevant part for this chapter is given in Table 13.

Table 13 Ranking for rhythmic base structure in Dutch

RT=TR ; FTBIN >> PARSE  $\sigma$  >> ALIGN-PRWD >> ALIGN- $\Sigma$

This grammar enables us to describe the longer rhythmic patterns in Figure 14.<sup>7</sup>

Figure 14 Possible rhythmic patterns in Dutch for phrases with more than four  $\sigma$ 's

- a. ( $\sigma$   $\sigma$ )  $\sigma$  ( $\sigma$   $\sigma$ )
- b. ( $\sigma$   $\sigma$ ) ( $\sigma$   $\sigma$ ) ( $\sigma$   $\sigma$ )
- c. ( $\sigma$   $\sigma$ )  $\sigma$  ( $\sigma$   $\sigma$ ) ( $\sigma$   $\sigma$ )
- d. ( $\sigma$   $\sigma$ ) ( $\sigma$   $\sigma$ ) ( $\sigma$   $\sigma$ ) ( $\sigma$   $\sigma$ )
- e. ( $\sigma$   $\sigma$ )  $\sigma$  ( $\sigma$   $\sigma$ ) ( $\sigma$   $\sigma$ ) ( $\sigma$   $\sigma$ )

In the OT grammar PARSE- $\sigma$  demands that syllables are parsed in a foot and FTBIN and RT=TR provide for the preference of trochaic feet. These constraints dominate PARSE- $\sigma$ , which results in an unparsed syllable for every phrase of an uneven number of syllables. ALIGN- $\Sigma$  demands that feet align with the right edge of the phrase. This constraint, however, would cause the unparsed syllable with an uneven number of syllables to align with the left edge of the phrase. This is avoided by a dominant alignment constraint ALIGN-PRWD which requires that the left edge of a phrase aligns with the left edge of a foot: ALIGN (PrWd, Left, Foot, Left). This constraint now ensures that a triplet pattern can only occur at the left edge of a phrase.

<sup>7</sup> For the sake of clearness we abstract from the influence of syllable weight here, and thus we start from a sequence of light syllables.

The ranking in Table 13 implies that the grammar does not enable us to describe the triplet pattern in Table 12b, because in that example the triplet pattern is carried through the whole phrase. The OT grammar will always prefer the structure in Figure 14e to that in Table 12b for a nine-syllable phrase. After all, the constraint ranking in Gilbers and Jansen (1996) leads where possible to a trochaic rhythmical basic pattern for Dutch phrases, like in *parallellogrammen* ‘parallelograms’, where s- and w-syllables nicely alternate. A compound like *bijstanduitkeringsgerechtigde* also gets a trochaic pattern – where allowed by morphological structure and difference in syllable weight (as in Table 12a).

However, there are languages in which the standard is a ternary rhythmic pattern. The dactyl pattern in Table 12b is a normal rhythmic pattern of prosodic words in languages like Estonian and Cayuvava: every s-syllable alternates with two w-syllables. Kager (1994) bases his analysis for patterns like those on *Weak Local Parsing theory* (Kager 1993; Hayes 1995): feet are at most binary and a ternary pattern is caused by an unparsed syllable between feet.<sup>8</sup> The constraint bringing about this effect is FOOT REPULSION: \*ΣΣ (avoid adjacent feet), it is a kind of OCP effect. FOOT REPULSION dominates PARSE-σ in languages like Estonian and Cayuvava. This constraint allows us to account for the alternative rhythmic structure of *bijstanduitkeringsgerechtigde* in Table 12b.

The basic assumption in OT is that constraints are universal. Because of the trochaic character of Dutch we have to assume that FOOT REPULSION is situated very low in the constraint ranking of Dutch, in any case it is dominated by FTBIN and PARSE-σ. In fast speech such a constraint ranking leads to a pile-up of many accents in a short period. Therefore, a longer interval between the accents is needed. In order to avoid clashes, beats are distributed over the phrase as evenly as possible. The distances between beats are enlarged, in order to avoid a staccato-like rhythm. Evidently, we are not dealing with phonetic compression here, but with restructuring, which can be obtained by assuming a special constraint ranking for fast speech (see Chapter 3 for a different analysis, however). In fast

<sup>8</sup> Contrary to the analysis of ternary patterns like (s w) <w>, Dresher and Lahiri (1991), Selkirk (1980), and Hewitt (1992) propose ternary feet: (s w w). Dresher and Lahiri propose an extra parameter, in order to achieve a branching head of a binary foot. The resulting ternary feet violate FTBIN.

speech FOOT REPULSION (\*ΣΣ) dominates PARSE-σ. The different constraint rankings for andante and allegro speech will then be as in Table 14.

Table 14 Rhythmic variability and speech rate

- a. andante ranking:  
Rt=Tr ; FtBin >> Parse-σ >> Align-PrWd >> Align-Σ >> \*ΣΣ
- b. allegro ranking:  
Rt=Tr ; FtBin >> \*ΣΣ >> Parse-σ >> Align-PrWd >> Align-Σ

As a general rule we will hypothesize in Chapter 3 that in allegro rankings markedness constraints and OCP-effects (\*CLASH, \*ΣΣ) are dominant, while in andante rankings CORRESPONDENCE constraints are far more important. Consider the different pronunciations of the word *tandpasta* ‘tooth paste’ in Table 15. In addition to clash avoidance there is a preference for assimilation in allegro styles. The functional explanation is that markedness constraints (ease of articulation) dominate correspondence constraints (ease of perception) in fast speech. Speaking fluently becomes more important and therefore unmarked structures are preferred from an articulatory point of view.

Table 15 Andante and allegro speech

	a. andante:	b. allegro:	
<i>tandpasta</i>	[tantpasta]	[tampəsta]	‘tooth paste’
	S    s    w	S    w    s	

Besides these kinds of rhythmic restructuring, we already spoke of the Phrasal Rule in Chapter 1. The preference for satisfaction of boundary-marking constraints (HAMMOCK) at the expense of violation of correspondence constraints (O-O CORR with the base of the adjective *tandheelkundige* ‘dentistry’) can also be seen in the data

in Figure 15, which all show secondary stress shifts to the lefthand phrase boundary in fast speech.

Figure 15 Rhythmic shifts to the left (Visch 1989)

a. andante						b. allegro					
					X						X
		X			X	X					X
X		X			X	X		X			X
X	X	X			X	X	X	X			X
X	X	X	X	X	X	X	X	X	X	X	X
<i>tandheelkundige dienst</i>						<i>tandheelkundige dienst</i> ‘dentistry service’					

data: *aardrijkskundig genootschap* ‘geographical association’,  
*zevensnarige luit* ‘seven-string lute’, *speciale aanbieding*  
‘special offer’

In Chapter 4 we will show that this phenomenon is a structure-marking phenomenon rather than a rhythmic one, and moreover a subject of accent rather than of rhythm. On the foot-level, accent is primarily a matter of rhythm. Stressed syllables are potential ‘anchoring sites’ for accent, because stress and accent are ideally aligned (Beckman 1986).

The difference between stress and accent can be compared to the distinction between the time-span reduction and the metrical analysis by Lerdahl and Jackendoff (1983). The accents are made up by the prominent chords in the hierarchy of the time-span reduction, which ideally align with metrically strong positions (*cf.* the time-span reduction preference rules (TSRPRs) in Chapter 1).

Thus far, we started from an account in which rhythmic adjustments are explained by re-ranking of constraints. We explained it this way, because a simple explanation of two constraints that switch in a ranking seems to be an elegant and straightforward explanation. However, this implies different grammars for different rhythmic structures, or for different speech rates or styles, which has some drawbacks. In Chapters 3 and 4 we will elaborate this account, but we will show that it is not a satisfactory account and we will introduce an alternative account for variation.

Most rhythmic adjustment phenomena we described in this section have not been based on systematic, objective testing. Cooper and Eady (1986) tested the eurhythmy rules empirically and they did not find systematic evidence for the stress adjustments reported in the literature. The observations of such phenomena are persistent, however, and the question is whether we should view these phenomena from a different perspective. What catches the eye is that all these alleged stress phenomena aim at regularity and rhythmic alternation. A different perspective could be that this regularity is not based on rhythm as syllable counting, but on rhythm as timing, or, following the section on rhythm and meter in this chapter, on the more abstract notion of meter. There seems to be some mechanism that requires stresses to fall on an ideal time interval from each other. This points to a dominant constraint 'METRONOME'. In the next section we will describe some more recent accounts of rhythm, on music as well as on speech, that take rhythmic timing into account. Chapter 3 will also lead us to the conclusion that rhythm perception is based on timing.

## 2.5. Rhythmic timing

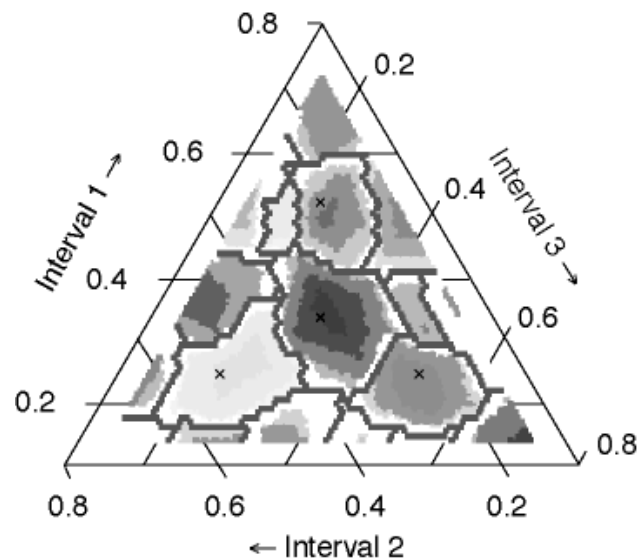
### 2.5.1. Restructured rhythm in music

A question in musical rhythm research is whether tempo changes imply relative invariance, or considerable deviation from proportional scaling. In other words, do the time intervals stay in the same ratio relative to each other when tempo changes, or are the relationships between rhythmic intervals distorted by the changed rate? Will all notes keep the same relative durations? Or will some notes get more prominence at the cost of other notes? This is a quite similar discussion to the one we will pick up in Chapter 3, where the notion of phonetic compression is the same as relative invariance and proportional scaling here, and rhythmic restructuring as the deviations from them.

Repp (1994, 1995) finds that in expressive timing, which expresses the phrasal structure of the performed music, only small deviations from proportionality occur. However, on the lower levels of rhythmic timing, considerable deviations from proportional

scaling are reported (Desain and Honing 1994). At different tempos, rhythm patterns are performed and perceived differently. Listeners do not perceive durations on a continuous scale. Instead, they recognize rhythmic categories that function as a reference relative to which the deviations in timing can be perceived. At faster rates, fewer categories of rhythm are perceived and more complex rhythms tend to drift to more simple rhythms. Honing (2002) and Desain and Honing (2003) define the simple rhythms and the related but more complex variants ‘surrounding’ them as ‘rhythm spaces’, as shown in Figure 16. The rhythm spaces are the categories in which in the listeners’ perception different rhythms tend to shift to the central rhythm. This is comparable to the way allophones are perceived as phonemes. Though all the allophones may be different, they group around the ‘central’ phoneme, and they are all perceived as one phoneme in the specific language.

Figure 16 Ternary plot of the rhythm space showing the areas of responses, centered around the points of the different rhythmic stimuli (Desain and Honing 2003)



Povel (1981) already found that subjects' reproductions of musical rhythms, although related to the stimulus ratios, are strongly distorted in the direction of a ratio of 1:2, 1:2 itself being reproduced most accurately. Collier and Wright (1995) also observe tendencies of reduction towards simple ratios at faster tempos, and towards contrast at slower tempos. The simple ratios to which the rhythms are drifted are never actually achieved, however (Repp 1998, Repp et al. 2002); the deviations from the optimal ratios probably lie within the same region of the rhythm space (Honing 2002). The preference for simple ratios implies that rhythmic patterns with many different note durations will be equalized to some extent in production, and are also perceived as more equal than they are.

The deviations of the rhythm are not proportionally scaled when tempo changes, at least in larger rhythmic groups. Repp et al. (2002) show that for three-note rhythms short intervals are timed quite accurately, whereas longer intervals show assimilation which increases as the tempo increases. Desain and Honing (1993, 1994) and Repp et al. (2002) conclude that relational invariance does not hold for rhythm at varying tempos: rhythm patterns are restructured. Tempo changes lead to a perceptual reorganization or regrouping of the rhythmic structure; long intervals turn into short intervals, and larger groups of events are formed. Sometimes it is even difficult to recognize the same rhythm at different tempos (Handel 1993). All results mentioned above point to the fact that rhythm is variable, and that it is strongly related to timing. This finding in music is the basis for our analyses in Chapter 3. In the next section we describe some accounts of speech rhythm as a timing phenomenon.

### 2.5.2. Timing in speech rhythm

In the perception of speech rhythm, not only the alternation of prominent and less prominent events are at stake, timing also plays a primary role, as we saw for music in the foregoing section. Port et al. (1998), Cummins (1997), Cummins and Port (1998), and Quené and Port (2003, 2005) integrate a mechanism known from extra-linguistic physical dynamical systems, self-entrainment (*cf.* Kelso 1995), into speech science. Self-entrainment means that the timing of repetitive motions by one oscillating system, like the pendulum of a clock, influences the motions of the other oscillator such that they fall into

simple temporal relationship with each other. That is, the oscillators tend to perform their motions in the same amount of time or in half (or double) the time, or in some other simple integer ratio of time. Speech rhythm is found to resemble such oscillating systems (Cummins 1997).

Cummins (1997) performed a couple of experiments with speech cycling tasks, and he found that the effect of harmonic timing is highly robust under conditions that encourage settling, not only in speech cycling tasks (repeating the same phrase), but also in chanting, poetry reading, singing while working together, etc. Speakers have a strong tendency to place the onsets of stressed syllables at temporal harmonic fractions of a metronome cycle, preferably at the largest harmonic fractions of 1: 1/2, 1/3, 2/3. These cycles are interpreted as isochronous series, if series can include silent beats as well (*cf.* Abercrombie 1965). Notice that this replicates the findings for musical rhythm (Povel 1981, Collier and Wright 1995, Honing 2002, Repp et al. 2002).

Quené and Port (2003) performed an experiment in which they aligned syllables of short phrases, such as *big for a duck*, with a metronome beat. They show that prominent vowel onsets are attracted to periodically spaced temporal locations, and this finding was formulated in the Equal Spacing Constraint. Vowel onsets are seen as the best approximation of P-centers (perceptual centers), the reference points for syllables on the basis of which listeners judge the timing of syllable or word sequences (Morton et al. 1976, Patel et al. 1999, Pompino-Marshall 1991).

Quené and Port also show that the Equal Spacing constraint gets stronger when the speaking rate increases, with an increasing number of stress shifts at faster speaking rates as a result. In fast speech, the number of beats is reduced, which appears to lead to a different prosodic structure containing less stressable positions, i.e. fewer feet. These feet in fast speech can consist of two or three syllables, or optionally of a single syllable supplemented with a silent beat: [*big for a*] [*duck* \_]. This finding suggests that produced stress shift is conditioned at least in part by the global rhythmic pattern of the utterance. Quené and Port (2003) performed some reaction time experiments on spoken word perception with more natural speech data. They find that rhythmical regularity helps in word recognition.



These findings are highly relevant for our investigations in secondary stress shifts and rhythmic restructuring in fast speech, therefore we will also take up this line of investigation in Chapter 3. The experiments in the papers discussed above, except for Quené and Port (2005), use quite unnatural speech tasks. In our own experiments we tried to find our evidence in a more natural experimental setup in order to get near-natural speech data. In the next chapter, we will test our hypothesis that speech rate, just as in music, influences the rhythmic structure.

## 2.6. Summary

In this chapter we gave an overview of some issues concerning rhythm in music and speech. In the first section we saw that rhythm is quite an omnipresent phenomenon in the world around us, which means it is not unique to speech or music. This chapter is concerned with rhythmic sound, i.e. patterned sound. We illustrated that rhythm can consist of either regular or irregular patterns, grouped into structures using accents to mark prominent elements.

The second section distinguished rhythm and meter. While rhythm consists of physical stimuli, meter – perfectly regular rhythm – is merely a psychological, subjective percept.

In the third section we showed that the linguistic rhythm-bearing units are the mora, the syllable, and the foot, and that rhythm is also influenced by stress patterns. On the basis of these rhythm units, a classification into three rhythm classes has been drawn up, into which languages of the world have been classified. Moreover, a nice parallel has been found between the musical rhythm of certain cultures and the rhythm classes of their languages.

In section 2.4 we described variability in speech rhythm, and we saw that speech rhythm has some ideal patterns to which it tends to conform. Fast speech often triggers triplet patterns. We introduced the constraint  $*\Sigma\Sigma$  (Kager 1994) in the constraint ranking to account for these triplet patterns.

In the last paragraph we discussed what happens to rhythm patterns under the influence of tempo changes. In speech as well as in music, the patterns tend to shift towards simpler ratios.

The last two subjects are the basis of the next chapter, in which we will investigate rhythmic changes under the influence of speech rate.

## Chapter 3

# The Influence of Speech Rate on the Perception of Rhythm Patterns

### 3.1. Introduction<sup>9</sup>

In Chapter 2 we introduced some processes of rhythmic variability. The topic of this chapter is how rhythmic variability in speech can be accounted for both phonologically and phonetically. Three lines of investigation are considered. The first is the question whether a higher speech rate leads just to 'phonetic compression', i.e. shortening and merging of vowels and consonants, with preservation of the phonological structure. As Schreuder and Gilbers (2004b) show, phonetic compression is evidently not the sole effect of fast speech. The second line of investigation is their claim that fast speech leads to adjustment of the phonological structure, and that the melodic content of a phonological domain is adjusted optionally when the speech rate increases, in order to obtain more eurhythmic patterns (Hayes 1984, Kager and Visch 1988, Van Zonneveld 1983). This claim is supported by the trained listener judgments of the outcomes of our experiment, as described in section 3.5.2. Conversely, the acoustic analyses lead to different insights and we will therefore investigate a third line (section 3.5.4), which concerns rhythmic timing in the perception of the listener, as indicated in Chapter 2.

In this chapter, we will first give the analysis based on the idea that clashes are avoided in allegro tempo. In Schreuder and Gilbers' proposal the restructuring phenomenon is explained by stating that every speech rate has its own preferred register, or - in terms of Optimality Theory (Prince and Smolensky 1993) - its own ranking of

---

<sup>9</sup> This chapter is an extension of Schreuder and Gilbers (2004b) and Schreuder and Gilbers (to appear). The results in those papers were based on a pilot experiment, while the current chapter concerns the full experimental data, with different outcomes and also a different conclusion.

constraints. This solution is controversial, because in standard Optimality Theory this would mean that each speech rate is described as a different language and that would be an odd description of such a minor difference. Therefore, we discuss three other models and we show that these models also face problems with respect to our data. We will give an alternative analysis, based on a variant of stochastic Optimality Theory (Boersma and Hayes 2001) which is called Simulated Annealing Optimality Theory (Bíró (to appear), Biro, Gilbers and Schreuder (to appear)).

As we pointed out in Chapter 1, our ultimate aim is to provide evidence for the assumption that all temporally-ordered behavior is structured similarly (*cf.* Liberman 1975). Gilbers and Schreuder (2002) show that Optimality Theory owes a lot to the constraint-based music theory of Lerdahl and Jackendoff (1983). Based on the great similarities between language and music we claim that musical knowledge can help in solving linguistic issues.

With regard to rhythmic restructuring, distances between beats are enlarged in both language and music, i.e. there appears to be more melodic content between beats. To illustrate this, we ran an experiment in which we elicited fast speech. As expected, speech rate plays an important role in the perception of rhythmic variability, as revealed by the auditory analyses of the data. However, as stated above, the acoustic analyses did not enable us to corroborate the claim of phonological restructuring. Therefore, we investigated a third possibility, namely that it is a perception rather than a production phenomenon. This perspective is a radically different approach from most work in laboratory phonology.

The chapter is organized as follows. In section 3.2 the data of the experiment is introduced. Section 3.3 addresses the rhythmic restructuring hypothesis in music and speech, the phonological framework of Optimality Theory, and the Simulated Annealing Optimality Theory analysis of the differences for *andante* and *allegro* speech. The method of the experiment is discussed in section 3.4 and the auditory and acoustic analyses plus the results and the phonological analysis follow in section 3.5. The conclusions and the perspectives of our analysis will be discussed in the final section.

### 3.2. Data

Following the literature on stress and rhythm (Hayes 1984, Kager and Visch 1988), we use the prevailing terminology of ‘stress shift’ for our rhythmic variability phenomena, although this term is not an optimal description, as we will show in this chapter. We will discuss three types of rhythmic variability in Dutch. The first type we will call “stress shifts to the right”, or in short “right shift”; the second “stress shifts to the left” or “left shift” and the third “beat reduction”. In the first type, as exemplified in *stúdietòelage* (S w s w w) ‘study grant’, we assume that this compound can be realized as *stúdietoelàge* (S w w s w) in allegro speech. *Perfèctioníst* (w s w S) is an example of “stress shift to the left” and we expect a realization *pèrfexioníst* (s w w S) in allegro speech. The last type does not concern a stress shift, but a stress reduction. In *zùidàfrikáans* (s s w S) ‘South African’, compounding of *zuid* and *afrikaans* results in a stress clash. In fast speech this clash is avoided by means of reducing the second beat: *zùidafrikáans* (s w w S). We used ten words of each type. Table 16 shows a selection of our data.

Table 16 Data<sup>10</sup>

Type 1: stress shift to the right (andante: S w s w w; allegro: S w w s w)

<i>stu die toe la ge</i>	‘study grant’
<i>weg werp aan ste ker</i>	‘disposable lighter’
<i>ka mer voor zit ter</i>	‘chairman of the House of Parliament’

Type 2: stress shift to the left (andante: w s w S; allegro: s w w S)

<i>per fec tio nist</i>	‘perfectionist’
<i>a me ri kaan</i>	‘American’
<i>pi ra te rij</i>	‘piracy’

Type 3: beat reduction (andante: s s w S; allegro: s w w S)

<i>zuid a fri kaans</i>	‘South African’
<i>schier mon nik oog</i>	‘name of an island’
<i>uit ge ve rij</i>	‘publishing company’

<sup>10</sup> Some examples from the experiment can be downloaded as mp3-files from <http://home.planet.nl/~schre537/sounds.htm> or [www.maartjeschreuder.nl](http://www.maartjeschreuder.nl).

In the *s s w s* structure Type 3 rhythms in e.g. *zuidafrikaans* (andante), *-a-* cannot be reduced, because generally reduction of a vowel to schwa is not possible in strong syllables. In fast speech, however, reduction seems to be possible. This would indicate the occurrence of restructuring: the second syllable fills a weak position. In a phonological account without restructuring in fast speech, this has no explanation. For this reason, we take musical rhythm theory into account: the reduction possibility can only be explained if the rhythm is simplified to a triplet, in which only the first note is strong. In the weak second position in the triplet reduction of the syllable *-a-* to schwa is possible (*cf.* Gilbers 1987).

The different rhythmic patterns are accounted for phonologically within the framework of Simulated Annealing OT.

### **3.3. Framework and phonological analysis**

#### **3.3.1. Rhythmic restructuring in music**

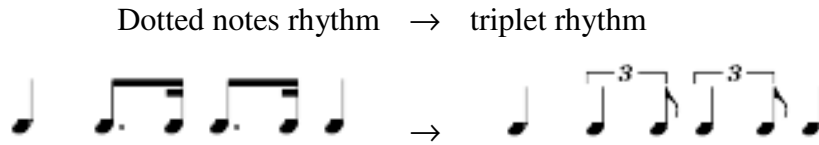
As mentioned in chapter 1, the mechanism of constraint interaction, the essential characteristic of OT, is also used in the generative theory of tonal music (Lerdahl and Jackendoff 1983). In both frameworks, constraint satisfaction determines grammaticality and in both frameworks the constraints are potentially conflicting and soft, which means violable. Violation, however, is only allowed if it leads to satisfaction of a more important, higher-ranked constraint. The great similarities between these theoretical frameworks make comparison and interdisciplinary research possible.

For example, restructuring rhythm patterns as a consequence of a higher playing rate is a very common phenomenon in music. Change of the tempo of a piece does not sound like a gramophone record played at the wrong speed, without changing the pitches; the structure changes in a higher tempo (*cf.* Repp 1990, 1995, Honing 2002, and Desain and Honing 2003). If in a music performance a piece is played at a different tempo, other structural levels become more important; for instance, at a lower tempo the tactus will shift to a lower level and the subdivisions of the beat will become more pronounced, or the other way around, in a higher tempo, some beats will get less prominence. An example of the last phenomenon is the

quadruple measure which is counted in four. When played in a faster tempo, the performer will sometimes choose to count it in two, hereby ‘de-accentuating’ counts 2 and 4. Normally it would be counted as ‘one - two - three - four, one - two - three - four’, with prominence on the first and the third count, with the first count as most prominent, yet now it becomes ‘one - ‘n’ - two - ‘n’ - one - ‘n’ - two’; the faster tempo moves the tactus to another level of the metrical structure, which gives the piece a different character. Because of the automaticity of this process, performers must sometimes be careful not to disrupt the specific character of a piece. Composers can prescribe whether the piece should be played in four or in two.

Musical experiments of Collier and Wright (1995) revealed noteworthy behavior related to different tempos, as we already pointed out in Chapter 2. In slower tempos, rhythmic contrast tended to be maintained or enhanced by differentiating between similar note onset intervals, whereas faster tempos resulted in reduction to more simple ratios between intervals. Similar findings are reported by Repp et al. (2002), who observed that rhythms are simplified towards simple ratios and that tempo has a strong effect on rhythmic performance in rhythms of more than two intervals. In Figure 17 we give an example of re-/misinterpretation of rhythm in accelerated or sloppy playing, which is well-known to be displayed by many musicians. This musical figure is the ultimate stumbling block for cellists and viola-players in the entrance requirements for the academy of music, as exemplified in the second movement (Andante con moto) of the famous Fifth Symphony in C minor opus 67 by Ludwig van Beethoven (Figure 18).

Figure 17 Rhythmic restructuring in music

Figure 18 Dotted notes rhythm in the second movement of Beethoven's 5<sup>th</sup>

Andante con moto

The “dotted notes rhythm” (left of the arrow) in Figure 17 is played as a triplet rhythm (right of the arrow). In the dotted notes rhythm the second note has a duration which is three times as long as the third, and in the triplet rhythm the second note is twice as long as the third. As shown by e.g. Repp et al. (2002), it is easier in fast playing to have equal durations between note onsets, or at least durations in simple ratios (*cf.* also Couper-Kuhlen 1993, Cummins and Port 1998, Port, Tajima and Cummins 1998, Quené and Port 2003). Clashes are thus avoided and one tries to distribute the notes over the measures as



evenly as possible, in spite of this implying a restructuring of the rhythmic pattern. To ensure that the beats do not come too close to each other in fast playing, the distances are enlarged, thus avoiding a staccato-like rhythm. In short, in fast tempos the musical equivalents of the Obligatory Contour Principle (OCP), a prohibition on adjacency of identical elements in language (McCarthy 1986), become more important.

### 3.3.2. Rhythmic restructuring in speech

If rhythmic restructuring in speech is a process that can be explained functionally by ease of articulation for the speaker - just as it is in music for the musician - the allegro patterns in all the different types of data in Table 16 would be caused by clash avoidance between main stress and secondary stress. The speaker would have a preference for beats that are more evenly distributed over the phrase.

Hence, Schreuder and Gilbers (2004b) described the different structures phonologically as a conflict between markedness constraints, such as FOOT REPULSION (\*ΣΣ) (Kager 1994), and OUTPUT - OUTPUT CORRESPONDENCE constraints (*cf.* Burzio 1998) within the framework of OT.<sup>11</sup> FOOT REPULSION prohibits adjacent feet and consequently prefers a structure in which feet are separated from each other by an unparsed syllable. This constraint is in conflict with PARSE-σ, which demands that every syllable is part of a foot. OUTPUT - OUTPUT CORRESPONDENCE compares the structure of a phonological word with the structure of its individual parts. For example, in a word such as *fototoestel* 'camera', OUTPUT - OUTPUT CORRESPONDENCE demands that the rhythmic structure of its part *tóestel* 'camera' with a stressed first syllable is reflected in the rhythmic structure of the output. In other words, OUTPUT - OUTPUT

---

<sup>11</sup> Elenbaas and Kager 1999 and Das 2001 replace \*ΣΣ with \*LAPSE, which interacts with the constraints ALL-FT-R and Parse-σ to account for ternary rhythm. To reach the same effect as \*ΣΣ, they had to extend the definition of \*LAPSE and to assume a gradient constraint ALL-FT-R. We choose to fall back on \*ΣΣ, because it stands for the avoidance of clashes on higher levels, as part of the OCP constraint family. It is also more generally applicable, e.g. for musical rhythm, where clashes on all levels are avoided.

CORRESPONDENCE prefers *fótotòestel*, with secondary stress on *toe*, to *fótotoestèl*, with secondary stress on *stel*.

Whereas the normal patterns in andante speech satisfy OUTPUT - OUTPUT CORRESPONDENCE, the preference for triplet patterns in fast speech is accounted for by Schreuder and Gilbers (2004b) by means of dominance of the markedness constraint, FOOT REPULSION, as illustrated in Table 17.<sup>12</sup>

Table 17 Rhythmic restructuring in language

a. ranking in andante speech:

constraints → <i>fótotoestel</i> candidates ↓	OUTPUT – OUTPUT CORRESPONDENCE	*ΣΣ	PARSE-σ
☞ ( <i>fóto</i> )( <i>tòestel</i> ) s w s w		*	
( <i>fóto</i> ) <i>toe</i> ( <i>stèl</i> ) s w w s	*!		*

b. ranking in allegro speech:

constraints → <i>fótotoestel</i> candidates ↓	*ΣΣ	OUTPUT – OUTPUT CORRESPONDENCE	PARSE-σ
( <i>fóto</i> )( <i>tòestel</i> ) s w s w	*!		
☞ ( <i>fóto</i> ) <i>toe</i> ( <i>stèl</i> ) s w w s		*	*

As mentioned in Chapter 2, Dutch is described as a trochaic language (Neijt and Zonneveld 1982). Table 17a shows a preference for an

<sup>12</sup> For reasons of clarity, we abstract from constraints such as FOOTBINARITY (FtBIN) and WEIGHT-TO-STRESS PRINCIPLE (avoid unstressed heavy syllables) in Table 17. Although these constraints play an important role in the Dutch stress system (cf. Gilbers and Jansen 1996), the conflict between OUTPUT-OUTPUT CORRESPONDENCE and FOOT REPULSION is essential for our present analysis.

alternating rhythm, conforming to the rhythms of the individual word parts. The dactyl pattern as preferred in Table 17b, however, is a very common rhythmic pattern of prosodic words in languages such as Estonian, Cayuvava, Chugach Alutiiq, Winnebago, and the Bangla-dialect Tripura Bangla<sup>13</sup>: every strong syllable alternates with two weak syllables (*cf.* Kager 1994, Das 2001 and references therein). Estonian, for instance, is a quantity-sensitive language, in which feet can consist of either one heavy syllable, two syllables of any structure, or three syllables, the last syllable being not heavy (Hint, 1973). Schreuder and Gilbers (2004b) assume that the rhythm grammar, i.e. constraint ranking, of Dutch allegro speech resembles the grammar of these languages.

This analysis proposed by Schreuder and Gilbers (2004b) has some weaknesses, however: can one claim that the Dutch native speaker suddenly switches to a different grammar above a certain speech rate? If so, why do we still observe alternations between the slow and the fast speech form? In the next subsection we will discuss three alternatives, extensions of the standard OT model which have been proposed to account for variation, and a fourth account proposed by Bíró (2005, to appear) and Bíró, Gilbers and Schreuder (to appear).

### 3.3.3. Alternative OT accounts of variation

The first two alternatives allow re-ranking within one grammar in a more elegant way (Anttila and Cho 1998, and Boersma and Hayes 1999), the third alternative allows some non-optimal candidates to emerge as alternative forms (Coetzee 2004). We will show that the Gradual Learning Algorithm (Boersma and Hayes 1999) as well as the theories of Anttila and Cho (1998) and Coetzee (2004) face difficulties, mainly because they cannot finetune the frequencies of variants in relation to the involved speech rate phenomenon. The fourth and most promising account of variability alternative is an elaboration of stochastic OT, using Simulated Annealing (Bíró

---

<sup>13</sup> Estonian is spoken in Estonia; Cayuvava is an extinct language of Bolivia; Chugach (Alutiiq) is a Yupik dialect (Eskimo-Aleut) spoken in the North of the USA and Siberia; Winnebago is a Mississippi Valley language, USA; and Tripura Bangla is a dialect of the Bengal language Bangla, India (Das 2001).

2005). This is a heuristic technique for finding the best solution of so-called NP-hard problems, i.e. problems that need much time and computational capacities, such as finding the optimal candidate in an OT system (Eisner 2000). We will show that Simulated Annealing provides us with the most adequate account of variable rhythm patterns with respect to speech rate (*cf.* Bíró (to appear), Bíró, Gilbers and Schreuder (to appear)).

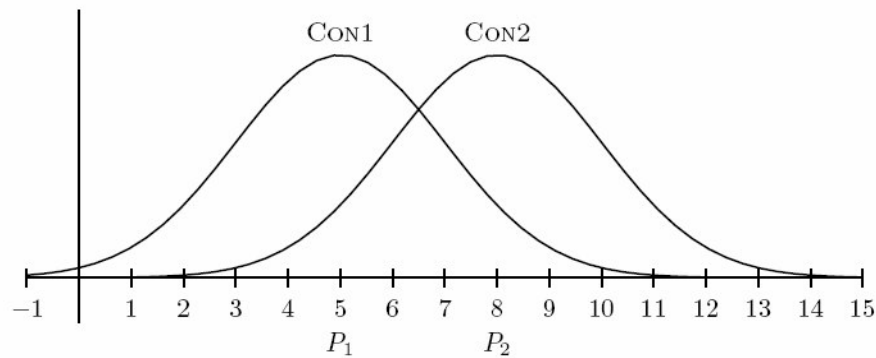
Anttila (1997) and Anttila and Cho's (1998) proposal uses a stratified hierarchy in which a subset of the constraints may be unranked in relation to each other. In their proposal the correspondence constraints and the markedness constraints should be placed on different randomly ranked strata, in order to deal with our data. A stratified hierarchy returns all candidates that are returned by at least one fully ranked hierarchy consistent with it. This way, we can include more hierarchies in the uniform description of a language, which may account for more candidates emerging on the surface. Furthermore, Anttila and Cho can make a prediction about the frequency of the varying forms. They hypothesize that each fully ranked hierarchy contained in a partial order has equal probability.

Applying Anttila (1997) and Anttila and Cho's (1998) proposal to our data would mean leaving \*ΣΣ and OUTPUT - OUTPUT CORRESPONDENCE unordered with respect to each other. Then, both outputs – (*stúdie*)(*tòelage*) and (*stúdie*)*toe*(*làge*) – are predicted to appear with a frequency of 50%. Consequently, this model is unable to account for the observed differences between andante and allegro speech, it only predicts the simultaneous existence of both forms. Moreover, it runs up against similar objections to those encountered by the proposal of Schreuder and Gilbers (2004b).

Boersma and Hayes's (1999) Stochastic Optimality Theory suggests a different solution to the problem of re-ranking constraints within a grammar. In this approach, the constraints are ranked on a continuous scale, and each constraint is assigned a real number defining their relative ranking. The original hierarchy is disturbed during evaluation by some random noise (Gaussian) with a standard deviation around zero, which may cause the constraints to overlap, possibly leading to constraint re-ranking (see Figure 19). The closer the two constraints on the real-valued scale and the bigger the noise (higher standard deviation), the higher the probability of the two

constraints being re-ranked. If the speech rate increases we expect more noise and more performance errors.

Figure 19 Constraints in Stochastic OT, with Gaussian noise with a standard deviation of 2.



By tuning the real numbers assigned to the constraints using the Gradual Learning Algorithm, this model may predict any frequency distribution. In Stochastic OT, the evaluation noise may cause the reranking of constraints that are close enough to each other. Our fast speech forms should then be seen as performance errors, which are the outcome of such a re-ranking. A low noise level then results in a production of the ‘right’ form, whereas increasing noise will increase the chance of ‘erroneous’ forms. This is a more elegant solution than both the models of Anttila (1997) and Anttila and Cho (1998), and Schreuder and Gilbers (2004b), because this approach leaves the grammar unchanged. The only thing that changes is in the evaluation noise.

In this proposal, the tableau in Table 17a presents the unperturbed grammar of Dutch secondary stress assignment, while the winner in Table 17b is the result of reranking the two constraints after adding noise, and must therefore be seen as a performance error. Suppose that constraint \* $\Sigma\Sigma$  is ranked only slightly lower than OUTPUT - OUTPUT CORRESPONDENCE, and the standard deviation (noise level) is relatively small, then the probability of the allegro speech form (*stúdie*)*toe(là)ge*) would be low. As the standard deviation grows,

however, and becomes comparable to the relative distance between the two constraints, the fast speech forms will emerge.

Some problems with this model arise, nonetheless. A theoretical one relates to the nature of the noise level. It could be postulated that increasing the speech rate corresponds to increasing the standard deviation defining the normal distribution of the evaluation noise. As speech rate grows, so does the standard deviation, causing the two constraints to be reranked more frequently, due to which the model returns the fast speech forms with a higher frequency. The question, however, is why the noise level would grow with speech rate. Future empirical research may be able to formulate a more exact connection between speech rate and noise level.

A second problem is that with the proposed constraints the frequency of the andante forms are predicted never to decrease below 50%, because the unperturbed ranking of OUTPUT - OUTPUT CORRESPONDENCE is higher than that of  $^*\Sigma\Sigma$ . This means that the chance of selecting points for which the ranking O-O CORR  $\gg$   $^*\Sigma\Sigma$  yields is always higher than the opposite ranking. That contradicts some of our empirical findings. In the results section of this chapter we will show that the fast speech form in some data exceeds the 50% in fast speech.

The most serious criticism is that for a given standard deviation of the noise level, the probability of the fast speech form is constant, independent of the input form. The emergence of the fast speech form is always the result of the same reranking. Whatever the type of words concerned (*cf.* Table 16), the probabilities will be identical. Our data show, however, very significant differences in frequency between the three rhythm types. For Stochastic OT this would mean that the standard deviation should not only be related to speech rate, but also to the input rhythm type.

Both approaches incorporating re-ranking into the grammar make some very strong predictions. For instance, whenever a number of constraints must be unranked with respect to each other in order to predict a given variation, then all other forms produced by other permutations of these forms must also be an attested variation, which may turn out to be problematic (*cf.* Bíró (to appear), Bíró, Gilbers and Schreuder (to appear)).

The third possible analysis of variation is to allow some sub-optimal output candidates to emerge as alternative forms. Coetzee

(2004) proposes a rank-ordering model in which the complete candidate set is harmonically ranked. In standard OT, only the optimal candidate will survive as output. In Coetzee's model the losing candidates are also ordered with respect to each other. In his view, the second-best candidate will be the second most frequently appearing variant of a certain form. Coetzee claims that candidates which are still in competition during the evaluation of the so-called critical cut-off point can be variants of the optimal candidate. Coetzee defines the position of the critical cut-off point as follows:

- (i) No candidate that is observed as a variant should be disfavored by any constraint ranked higher than the cut-off.
- (ii) All candidates that are not observed as variants should be disfavored by at least one constraint ranked higher than the cut-off (Coetzee 2004: 167).

His prediction is that whenever the third best candidate is observed as an alternative form, then the second best must also appear in the language.

This account is in itself an elegant solution to variation. Again, however, it is not the optimal analysis to account for our variation data, as Coetzee says nothing about the frequencies of the alternatives. The proposal attempts only to give qualitative, or relative, predictions about frequencies of alternating forms, no quantitative, or absolute predictions. In the results section of this chapter, we will show that some of our fast speech data are only characterized by a shift in the observed frequencies, not by relative occurrence per se. An account of variation should be able to deal with this.

Consequences of Coetzee's model include that whenever the third best candidate is observed as an alternative form, then the second best must also appear in the language. Furthermore, if the fourth best candidate is defeated at the same constraint as the third one, then the fourth one should also be an attested alternation form, otherwise we cannot identify the critical constraint. Bíró, Gilbers and Schreuder (to appear) also show with progressive voice assimilation data that sometimes a candidate is predicted to be a possible variant by Coetzee's model, while this candidate does not in fact occur. What is more, as the attested fast speech form violates the highest ranked

constraint, as in the tableaux in Table 17, the cut-off point must be set at the top of the hierarchy, wrongly predicting all candidates to surface in the language if we gave more candidates.

An alternative approach is proposed by B  r   (2005 / to appear). B  r  's Simulated Annealing Optimality Theory, although very different from it, resembles Coetzee's theory in that Simulated Annealing also allows non-optimal candidates to emerge as alternating forms. Simulated annealing, also called 'stochastic gradient ascent', is a model originating in statistical physics (Kirkpatrick et al. 1983). It is a heuristic technique for finding a good solution to an optimization problem. In an optimization problem, one searches for the element in a set that minimizes or maximizes a certain function. The goal of heuristic techniques is to return some solution to the problem quickly, although you cannot be sure whether you will really find the optimal solution. Still, the solution returned is 'near-optimal' (Reeves 1995). In many cases, it is not feasible to run an algorithm that guarantees finding the best solution, yet it is satisfactory to find a relatively good solution.

Let us illustrate this with a metaphor. A very simple heuristic optimization algorithm is 'gradient ascent', or 'hill-climbing-in-a-fog' (B  r   to appear, B  r  , Gilbers, and Schreuder to appear). Imagine someone wants to find the highest point of a landscape, while he can't see anything because of a dense fog. He randomly walks in the country, but the rule is never to move downhill. Clearly, he will soon reach the top of some hill: a *local* maximum, a position that is higher than any of its neighboring positions. Nothing guarantees, however, that he has reached the highest point of the whole search space, the *global* maximum. In fact, it is very likely that he has got stuck in a non-global, local maximum. This optimization algorithm is called *gradient ascent*.

Let us now change the rule, and introduce Simulated Annealing: it is now also permissible to move downhill. Before each step, one can randomly choose a direction, horizontal, or uphill, or downhill. If the neighbor chosen is higher or equally high, the random walker certainly moves there, whereas if the neighbor chosen is located lower, then the probability of moving to that point is smaller than one. The steeper the step downwards, the smaller the probability of



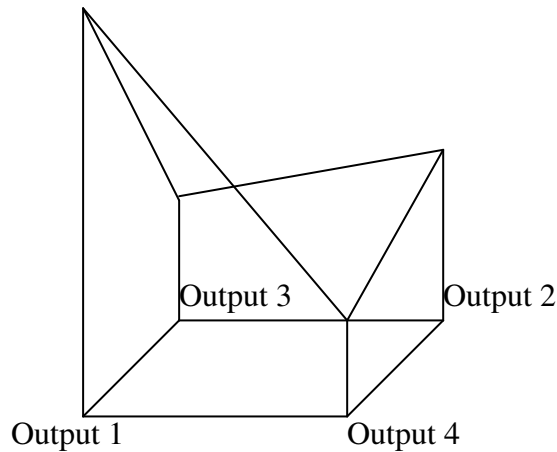
that step.<sup>14</sup> He does not know, however, whether the maximum he has reached is the *global* maximum or a *local* one. The walker can decide he is satisfied with the top he gets to first, but if he is eager to reach the highest mountain, he will try harder, and longer. He will be more precise. To put it in terms of Simulated Annealing: the parameter "temperature" is reduced in a number of steps (iterations) from its maximum value to its minimum value, and the algorithm finally returns the position of the random walker when temperature reaches its minimum. If the walker is given a high number of moves to perform (a high number of iterations; that is, temperature is reduced in small steps), then it is more likely that he will reach the global maximum. With less iterations (corresponding to temperature being reduced more quickly, in bigger steps), however, the chance of ending up in a local maximum is higher.

This search strategy can be applied to Optimality Theory. It enriches the candidate set with a neighborhood structure. The horizontal structure is made up of points, representing the output candidates. It may include an infinite set of candidates. Neighboring candidates are candidates that differ on a single aspect, structurally, segmentally, prosodically, etc. To this horizontal topology, the OT constraint ranking adds a vertical geometry, with peaks and valleys, and steep and shallow slopes. The global peak, or the global optimum, is the optimal output candidate of the OT constraint ranking. The goal of Simulated Annealing, is to find this global optimum. The topology also has other local optima, which may emerge as alternative forms if simulated annealing fails to find the global optimum. A candidate is a global or local optimum if and only if it is better than its immediate neighbors, with respect to the given hierarchy. Figure 20 illustrates a hypothetical neighborhood structure.

---

<sup>14</sup> Furthermore, the probability of moving one unit downhill decreases because of a parameter called "temperature" which decreases during the algorithm (hence the name "Simulated Annealing").

Figure 20 Hypothetical neighborhood structure for Simulated Annealing OT



When this neighborhood structure, Output 1 is the global optimum, as decided by a hypothetical constraint ranking. Output 2 is a local optimum, and may appear as a variant. Output 3 is also located on a relatively high point in the landscape, but is lies in between two higher points, and will therefore never be a local optimum.

If the neighborhood structure is defined, the simulation must be run. Table 18 gives the algorithm of Simulated Annealing. The simulation starts searching the neighborhood structure each time from a different search point, or output candidate, looking for the global optimum. This simulation is run an  $x$  number of times, e.g. a thousand times. The parameter 't\_step' ('t' for 'temperature') in the algorithm defines the size of the steps between 't\_max' and 't\_min'. If t\_step is set at e.g. 0.5, the steps are big and the simulation is run fast, while t\_step of e.g. 0.1 gives a more precise simulation. A fast simulation will thus be less precise in finding the global optimum, and may find a local optimum instead. From whatever point the simulation starts, the global optimum will be returned most often with a slower simulation, and the local optima will appear in relatively stable ratios over different simulations. In a fast simulation, local optima will be found more frequently and may appear to

emerge more often than the global one. A fast simulation may simulate fast speech, which is also less careful than moderate speech, and therefore may a number of times result in sub-optimal forms.

Running the simulation and tuning its parameters thus may or may not reproduce the observed data by returning the global and local optima with the expected frequency. Bíró (to appear) tested his algorithm with our fast speech data, and the results will appear in section 3.5.3. If Simulated Annealing can indeed predict the right frequency distributions of the possible output forms, this will be the best solution to our data.

Table 18 The algorithm of Simulated Annealing Optimality Theory (for a more detailed description of the model, see Bíró (to appear)).

```

ALGORITHM: Simulated Annealing for Optimality Theory
Parameters: w_init, K_max, K_min, K_step, t_max, t_min, t_step
# t_step: number of iterations / speed of simulation
w <-- w_init ;
for K = K_max to K_min step K_step
  for t = t_max to t_min step t_step
    choose random w' in neighborhood(w) ;
    calculate < C , d > = ||H(w')-H(w)|| ;
    if d <= 0 then    w <-- w'
    else
      w <-- w' with probability
      P(C,d;K,t) = 1      , if C < K
                  = exp(-d/t) , if C = K
                  = 0      , if C > K
  end-for
end-for
return w

```

Another promising objective of Simulated Annealing is concerned with a problematic issue, pointed out by Keller and Asudeh (2001) as a general problem in mainstream OT: if an output candidate is harmonically bound by the two alternative forms (Samek-Lodovici

and Prince 1999), it is an eternal loser and is predicted not to emerge as an alternating form. Bíró (to appear), and Bíró, Gilbers and Schreuder (to appear) show that sometimes such a harmonically bound candidate does appear as a variant. With Simulated Annealing such harmonically bound output forms can emerge in our fast speech data, as shown by Bíró (to appear). Reranking theories cannot account for these variants, while, theoretically, Coetzee's model may also allow harmonically bound forms to emerge in languages.

In the next section we will explore whether we can find empirical evidence for rhythmic restructuring in fast speech, and whether Simulated Annealing would make the right predictions about the frequency distributions.

### **3.4. Method**

#### **3.4.1. Subjects and task design**

To find out whether people indeed prefer triplet patterns in allegro speech, we did an experiment in which we tried to elicit fast speech. Twenty-five native speakers of Dutch (twelve men and thirteen women aged 11 to 42) participated in a multiple-choice quiz in which they competed with each other in answering thirty simple questions as quickly as possible. In this way, we expected them to speak fast without concentrating too much on their own speech.

The results of a pilot experiment had revealed that the fast subjects displayed more variability in their rhythmic patterns due to tempo than the slower subjects did, which means that their andante and allegro utterances had different rhythmic patterns in more instances. In order to see whether this observation would hold for the wider range of fast speakers, we decided to look for subjects who were known to speak very fast. We asked colleagues and friends if they knew such people. Of course, everyone knows notoriously fast-speaking people. The potential subjects were only told they were very suitable to participate in our experiment. Curious as they were why they would be suitable, they all immediately agreed to participate. We repeated the same experiment with the twenty-five fast-speaking subjects and thirty test words. In Table 19 one of the quiz items is depicted.

Table 19 Quiz item

Q	<i>President Bush is een typische</i> 'President Bush is a typical' ...	
A1	<i>intellectueel</i>	'intellectual'
A2	<i>amerikaan</i>	'American'
A3	<i>taalkundige</i>	'linguist'

### 3.4.2. Analysis methods

We categorized the obtained data as allegro speech (*cf.* section 3.5.1). As a second task the subjects were asked to read out at a normal speaking rate the answers embedded in the sentence *ik spreek nu het woord ... uit* 'now I pronounce the word ...'. This normal speaking rate generally means that the subjects will produce the words at a rate of approximately 180 words per minute, which we categorize as andante speech. All data were recorded on DAT tape (DAT recorder: Sony DTC-57ES; microphone: Sennheiser MKH 40 (mono); mixer: Eela audio S102) in a soundproof studio and digitalized and normalized in CoolEdit in order to improve the signal-noise (S/N) ratio. Normalizing to 100% yields an S/N ratio approaching 0 dB. This resulted in a data set of about 1500 words, of which half was in allegro tempo, the other half in andante tempo.

Five trained listeners judged the data auditorily and indicated where they perceived secondary stress. After this auditory analysis the data were phonetically analyzed in PRAAT (Boersma and Weenink 1992-2006). We compared the andante and allegro data by measuring duration, pitch, intensity, spectral balance and rhythmic timing, as described below (e.g. Sluijter 1995, Couper-Kuhlen 1993, Cummins and Port 1998, Quené and Port 2003). Sluijter claims that duration and spectral balance are the main correlates of primary stress. In our experiment, we are concerned with secondary stress.

For the duration measurements, the rhymes of the relevant syllables were observed. For example, in the allegro style answer A2 *amerikaan* in Table 19, we measured the first two rhymes and compared the values in seconds with the values for the same rhymes at the andante rate. In order to make this comparison valid, we equalized the total durations of both realizations by multiplying the duration of the allegro with a so-called 'acceleration factor', i.e. the duration of the andante version divided by the duration of the allegro

version. According to Eefting and Rietveld (1989) and Rietveld and Van Heuven (1997), the just noticeable difference for duration is 4,5%. If the difference in duration between the andante and the allegro realization does not exceed this threshold, we consider the realizations to be examples of the same speech rate and neglect them for further analysis.

For the pitch measurements, we took the value in Hz in the middle of the vowel. The just noticeable difference for pitch is 2,5% ('t Hart et al. 1990). For the intensity measurements, we registered the mean value in dB of the whole syllable.

The next parameter we considered concerns spectral balance. Sluijter (1995) claims that the spectral balance of the vowel of a stressed syllable is characterized by more perceived loudness in the higher frequency region, because of the changes in the source spectrum due to a more pulse-like shape of the glottal waveform. The vocal effort, which is used for stress, generates a strongly asymmetrical glottal pulse. As a result of the shortened closing phase, there is an increase of intensity around the four formants in the frequency region above 500 Hz. Following Sluijter (1995) we compared the differences in intensity of the higher and lower frequencies of the relevant syllables in both tempos.

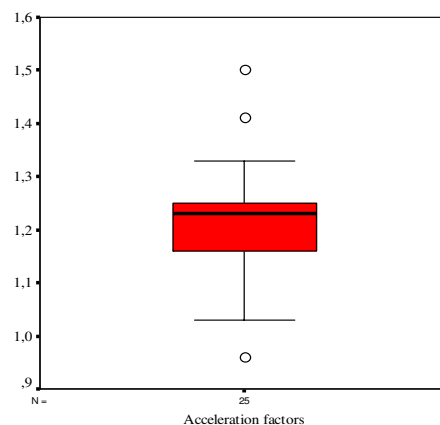
Finally, we considered rhythmic timing. The idea is that the beats in speech are separated from each other at an approximately equal distance independent of the speech rate. In other words, a speaker more or less follows an imaginary metronome. If he/she speaks faster, more melodic content will be placed between beats, which results in a shift of secondary stress. This hypothesis will be confirmed if the distance between the stressed syllables in the andante realization of an item, e.g. *stu* and *toe* in *studietoelage*, approximates the distance between the stressed syllables in the allegro realization of the same item, e.g. *stu* and *la*. If the quotient of the andante beat interval duration divided by the allegro beat interval duration approximates 1, we expect perceived restructuring.

### 3.5. Results

#### 3.5.1. Evaluating the task design

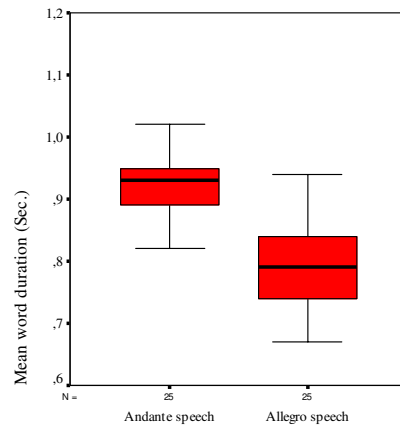
As for the pilot experiment, before looking into the auditory results, we first investigated whether the quiz design was successful: did the subjects speak faster in the quiz task than when we asked them to speak at a moderate speaking rate? We calculated the acceleration factors by dividing the mean total word durations of the andante words per subject by the durations of the allegro words. The boxplot in Figure 21 shows that all subjects but one had an acceleration factor above 1, which means they accelerated. And most of them accelerated quite strongly, which is indicated by the fact that their acceleration factors lie above 1.1. The mean is 1.2, the median 1.23. There are three outliers, indicated by the small circles in the boxplot. Two of these subjects have extremely high acceleration factors, which means they made a huge difference between their allegro and their andante speech. One subject has an acceleration factor below 1, which means his allegro speech was in fact slower than his andante speech. For this subject the quiz design obviously did not work: after coming up with the answer quickly, he spoke carefully and with clear articulation.

Figure 21 Boxplot of the acceleration factors: andante word durations divided by allegro word durations



The acceleration factors do not tell us whether the speakers spoke really fast, they only give insight into the differences in durations of the andante and the allegro words. The real durations of the andante and allegro words are shown in the boxplots in Figure 22, calculated for each subject. The subjects with the higher acceleration factors were also found among the speakers with the shortest allegro word durations, although there was no one-to-one mapping of highest acceleration factor to fastest allegro words. The mean duration of the andante words is 0.936, the median 0.93; the mean and median of the allegro words are 0.79. This means a difference between the means of andante and allegro word durations of 146 milliseconds, which is highly significant ( $t(24) = 8.439$ ,  $p < 0.001$ ).

Figure 22 Boxplots of the mean durations of andante and allegro words



From these data we can conclude that the quiz design was successful in eliciting fast speech.







### 3.5.2. Auditory analysis

The data of the full experiment was judged by five trained listeners, who either decided on which syllable in the words they perceived secondary stress, or they could choose which of the rhythms in the



two columns in Table 20 was more like the rhythm of the word, especially in the Beat Reduction cases.

Table 20 The possible rhythms of the test words

	A (Correspondence)	B (Restructured)
a. Right Shift	 <i>stu die toe la ge</i>	 <i>stu die toe la ge</i>
b. Left Shift	 <i>per fec tio nist</i>	 <i>per fec tio nist</i>
c. Beat Reduction	 <i>zuid a fri kaans</i>	 <i>zuid a fri kaans</i>

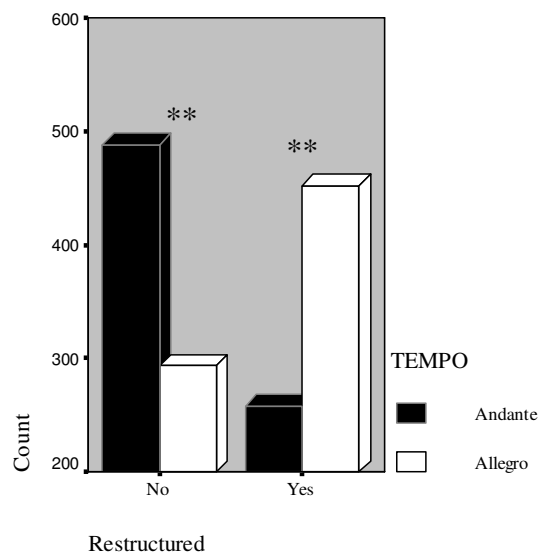
If the majority of these judgments indicated a restructuring as regards the correspondence pattern, we assigned this word 1 (yes), otherwise 0 (no). These judgments were analyzed with the help of a Pearson Chi-Square test, of which the results are shown in Figure 23. The graph clearly shows that the number of restructurings depends on the tempo. In andante tempo, 488 items are not restructured and conform to the correspondence pattern, while in allegro speech 452 items are rhythmically restructured, as can be seen in the cross tabulation of Figure 23. These differences are highly significant ( $\chi^2(1) = 101.695$ ,  $p < 0.001$ ). Therefore this proves the relation between speech rate and rhythmic restructuring. Moreover, when we take the word duration measurements and the acceleration factors into account, a Multivariate Analysis of Variance (MANOVA) shows a highly significant difference between the word durations and acceleration factors of words which were perceived as rhythmically restructured

and those which were not, as can be seen in Table 21 (total durations:  $F(1) = 5.908$ ,  $p < 0.001$ ; acceleration factors:  $F(1) = 2.156$ ,  $p < 0.001$ ).

Table 21 MANOVA Descriptive Statistics of the durations and acceleration factors of words perceived as either restructured or not restructured

	Restructured	Mean	Std. Deviation	N
Total duration	No	.924	.1937	779
	Yes	.798	.1535	710
Acceleration factors	No	1.07	.187	779
	Yes	1.14	.239	710

Figure 23 Numbers of perceived restructurings in the andante and allegro data



$$\chi^2 (df1) = 101,695; p < 0,001$$

(Table of Figure 23) Restructured \* TEMPO Cross tabulation

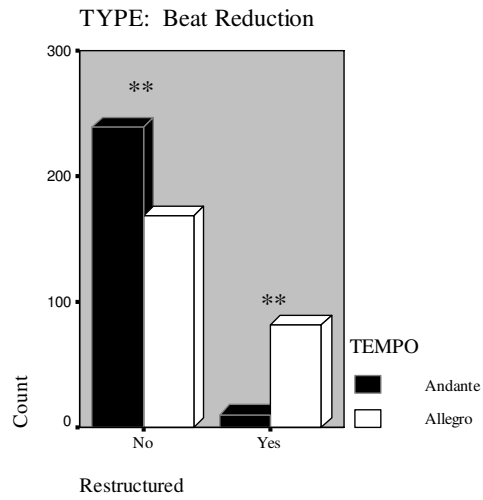
Count

		TEMPO		Total
		Andante	Allegro	
Restructured	No	488	293	781
	Yes	258	452	710

#### 3.5.2.1. Between-type variation

We further investigated how the three separate rhythmic types contributed to the result. Therefore, we split the data by rhythmic type, i.e. Beat Reduction (BR), Left Shift (LS), and Right Shift (RS). In Figure 24a,b,c we see clear differences between the types; the Left Shifts deviate most from the other two types and from the overall pattern in that they have a strong preference for restructuring. Still, all three types show significant differences between andante and allegro, if we take into account the fact that there is a bias for restructuring in the Left Shift cases, or for correspondence in the other two types.

Figure 24a Data split by Rhythmic Type



$$\chi^2(1) = 67.781, p < 0.001$$

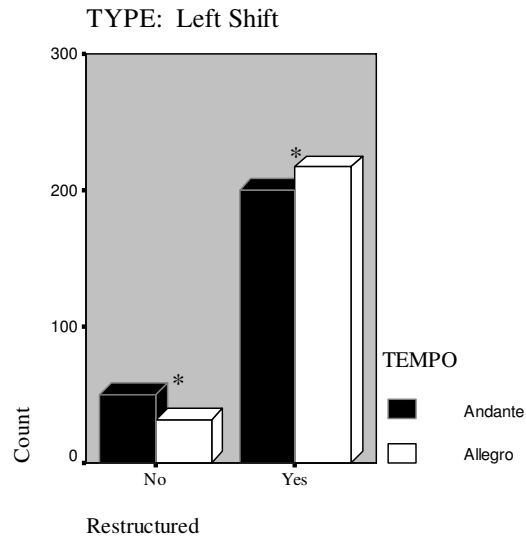
(Table of Figure 24a) shifted \* TEMPO Crosstabulation

Count

		TEMPO		Total
		1	2	
Shifted	0	239	168	407
	1	10	81	91

TYPE = BR

Figure 24b



$$\chi^2 (1) = 4.642, p < 0.05$$

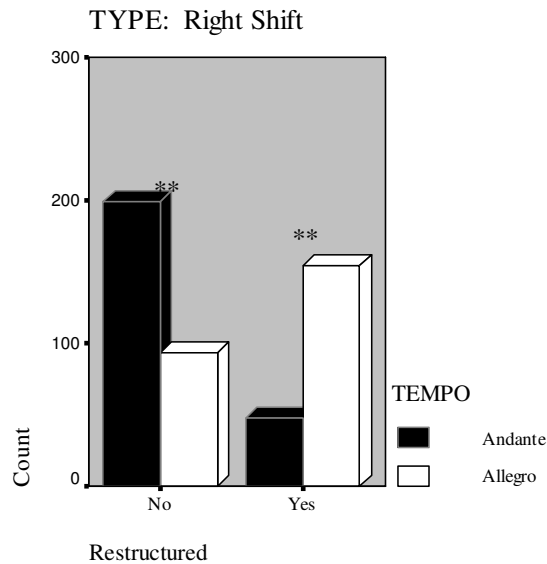
(Table of Figure 24b) shifted \* TEMPO Crosstabulation

Count

		TEMPO		Total
		1	2	
shifted	0	50	32	82
	1	200	217	417

TYPE = LS

Figure 24c



$$\chi^2 (1) = 94.103, p < 0.001$$

(Table of Figure 24c) shifted \* TEMPO Crosstabulation

Count

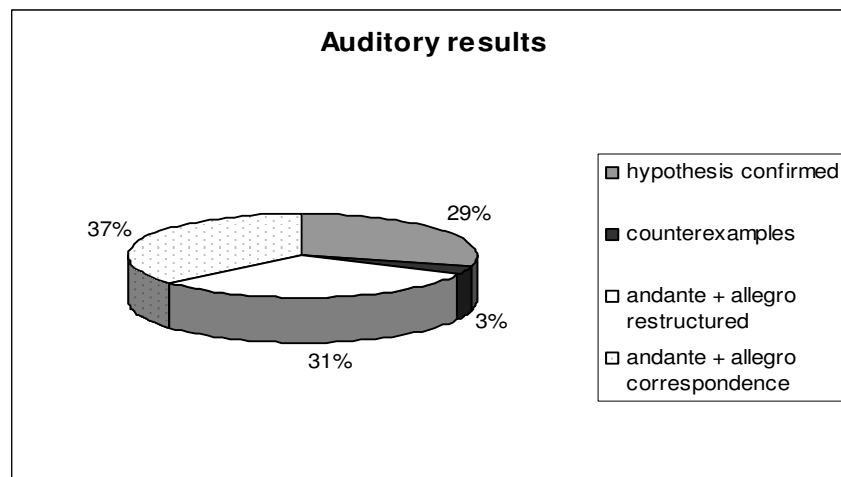
		TEMPO		Total
		1	2	
shifted	0	199	93	292
	1	48	154	202

TYPE = RS

Obviously, the preference for restructuring the rhythmic pattern in allegro speech is not an absolute preference. Sometimes restructuring does not take place in allegro speech, but on the other hand restructured patterns also show up in andante speech. The frequencies of the patterns are clearly dependent on speech rate, however. Some items were realized with the same rhythmic pattern irrespective of the tempo. Therefore, besides looking at the rhythmic

structures of all andante and all allegro data, we were also interested in what the data look like if we look at pairs of the same word by the same speaker, in andante and allegro tempo. In other words, how many times does a word in andante tempo show the correspondence pattern, while it shows the restructured pattern in allegro tempo? Instances of this pattern would strengthen the confirmation of our hypothesis. Furthermore, how many times is it the other way around, as counterexamples? The graphs in Figure 25a,b demonstrate that most of the time the words show the same pattern in both tempos, either both with a correspondence pattern or both with a restructured pattern. In almost one third of the instances, however, the hypothesis is confirmed, while the number of counterexamples is marginal.

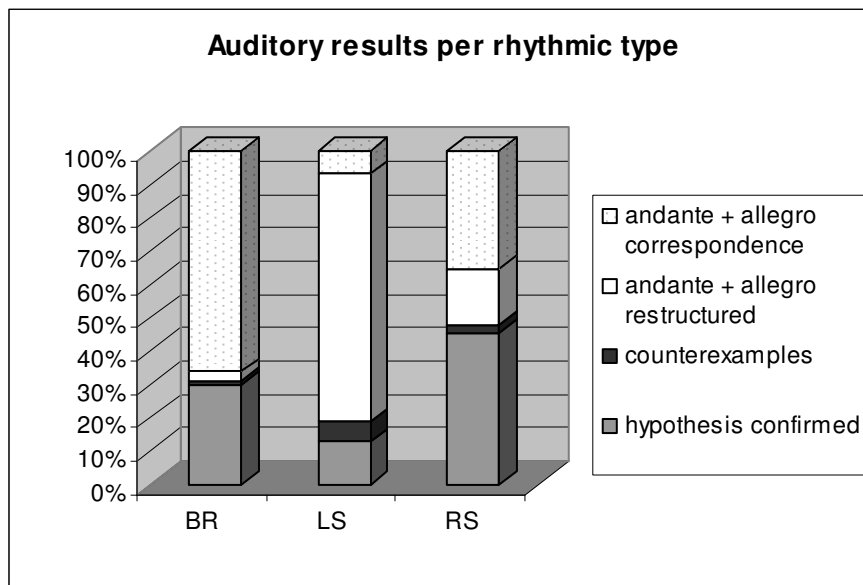
Figure 25a Pairs of the same word in andante and allegro



In Figure 25b we look at the same pairs, now split by rhythmic type. We see that the three types behave differently, as we saw in the graphs in Figure 24. Again, the Left Shift words show a strong preference for restructured patterns in both tempos, and a relatively high amount of counterexamples. The Beat Reduction words mostly conform to the correspondence pattern, whereas the Right Shift words are found to confirm our hypothesis in a majority of cases. However, the following observation holds for all three types: if the

andante and allegro patterns are different, they differ mostly in the direction of our hypothesis: andante words with a correspondence pattern, allegro words with a restructured pattern. The overall number of counterexamples is low, except maybe for the Left Shift words.

Figure 25b Pairs of the same word in andante and allegro, per rhythmic type



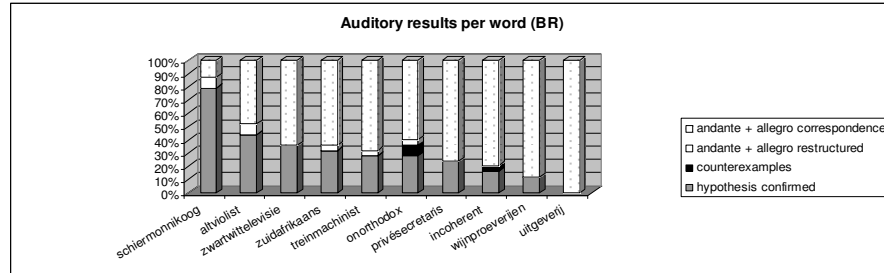
### 3.5.2.2. *Between-item variation*

If we separate the results further and look at the behavior of the individual words in Figure 26a,b,c, we see that in the Left Shift words (Figure 26b) many individual words are responsible for the high number of counterexamples. Some of the Left Shift words, *demoniseren* ‘demonize’, *specialiteit* ‘speciality’, *legaliseren* ‘legalize’, are always restructured, which may point to a certain degree of lexicalization with shifted secondary stress in this type of word; one of the Beat Reduction words, *uitgeverij* ‘publishing company’ is never restructured. Another Beat Reduction word, *Schiermonnikoog* ‘name of an island’, has a very strong preference

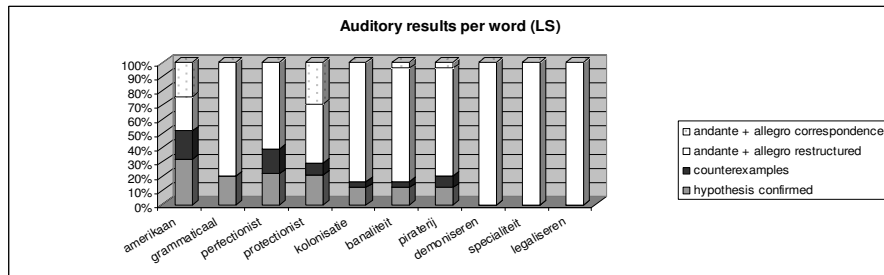


for correspondence in andante and restructuring in allegro tempo, the expected pattern according to our hypothesis. This can also be said of three of the Right Shift words, *winkelopheffing* ‘shop closure’, *trimesterindeling* ‘trimester distribution’, and *zenderinstelling* ‘channel tuning’.

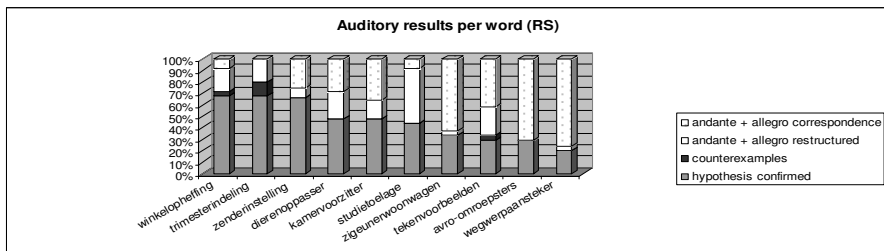
Figure 26 Pairs of the same word in andante and allegro, per word



a. Beat Reduction



b. Left Shift



c. Right Shift

Interestingly, these three Right Shift words are all nominalized verbs ending in the morpheme *-ing*. There are also four other nominalized verbs in the Right Shift type. These end in *-er(s)*. Two of those also score quite high, the other two, conversely, score lowest. In the other two rhythmic types no such observations can be made. On the basis of these data we cannot decide whether this would be more than coincidence.

Possibly, the syllable structure also plays a role; open syllables seem to lose stress somewhat more easily than closed ones. This is not clearly the case, however. It obviously depends on the rhythmic type: the 'left shift' words are far more often subject to rhythmic restructuring than the other two types. Most of these same words also have open syllables in the originally secondary-stressed syllable positions, but this is not the case for the often restructured words of the other rhythmic types.

What does seem to play a role is the morphological structure of the words. The types RS and BR are compounds, whereas the LS-type words are single derived words. The compounds have much more resistance to restructuring. Rhythmic restructuring of these words means ignoring or forgetting the morphological structure. The fact that an important part of these words is restructured, suggests that, in fast speech, rhythm does not depend on morphological structure, or one might say that a speaker makes use of a 'different lexicon', in which case these words are not compounds, but single words. In this last option, on the other hand, one would no longer expect these speakers to draw a distinction between these compound types and the 'left shifts', while they certainly do (*cf.* Figure 25b).

What is more, foot type seems to have its influence: the LS words start with iambic feet, while Dutch has a preference for trochees. This influence appears to be stronger than CORRESPONDENCE.

### 3.5.2.3. *Between-subject variation*

The subjects show quite a lot of variation; still their overall patterns are mostly similar. More importantly, the faster speakers were also those who differentiated rhythmically between the words in *andante* and *allegro tempo*, which means they not only displayed a greater difference in word durations in *andante* and *allegro* speech, but also more variability in their speech patterns due to tempo than the less

fast subjects do. This observation strengthens our claim that restructuring relates to speech rate.

The five trained listeners who judged our data had high mutual agreement in about 80% of the cases. In the other 20% agreement was low. We must say the listeners found it sometimes very hard to decide where they heard secondary stress. Especially the Beat Reduction data appeared to be very hard to judge. In some cases they couldn't decide. Sometimes an item was ambiguous; there seemed for example to be a pitch accent on the first syllable, while the listeners perceived a longer duration on the second syllable, or the syllables sounded equally strong. One listener remarked that some subjects often produced no secondary stress at all, in his perception.

### 3.5.3. Phonological analysis: Simulated Annealing

Finally, we will compare our outcomes with the outcomes of the Simulated Annealing by Bíró (2005). The relevant constraint ranking for the simulation is given in Table 22.

Table 22 The constraint ranking used in Simulated Annealing

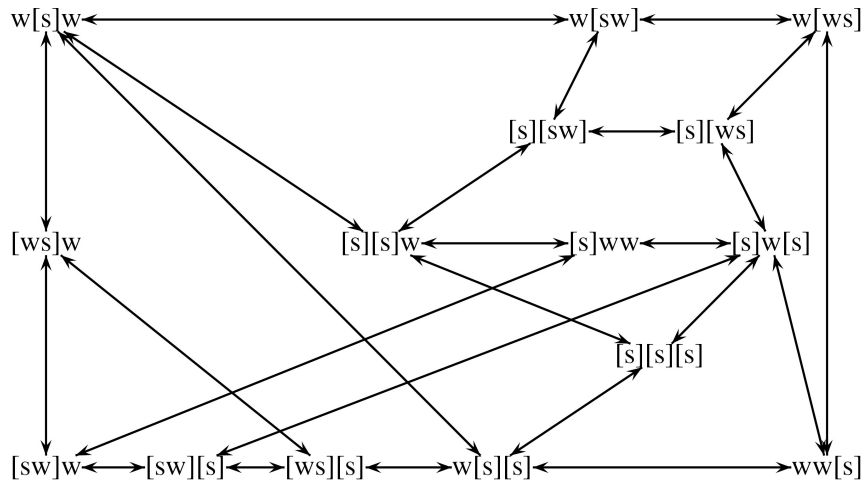
FTT=TR >> AFFIX-STRESS >> O-O CORR >> \*ΣΣ >> PARSE-σ

The output candidates are put in a neighborhood structure, shown in Figure 27. This figure shows which output candidates are included in the simulation, and which candidate forms are neighbors in the neighborhood structure. How the simulation works is described above in section 3.3.3. The vertical component of the landscape is based on the constraint ranking in Table 22, but is too complicated to show in a three dimensional picture. For the three rhythmic types, the vertical geometries are different, because of the presence of O-O CORR in the constraint ranking, and because the three types correspond to the different structures of the individual word parts.<sup>15</sup> Each rhythmic type thus has different local and global optima, and slopes, and will therefore show different frequency distributions. The

<sup>15</sup> Bíró slightly redefines O-O CORR, which enables the simulation to count the number of violation marks assigned by this constraint to any candidate.

simulation was run several times for each type, and the frequencies of the output candidates are shown in Figure 28, where they are compared to the observed frequencies in our data.

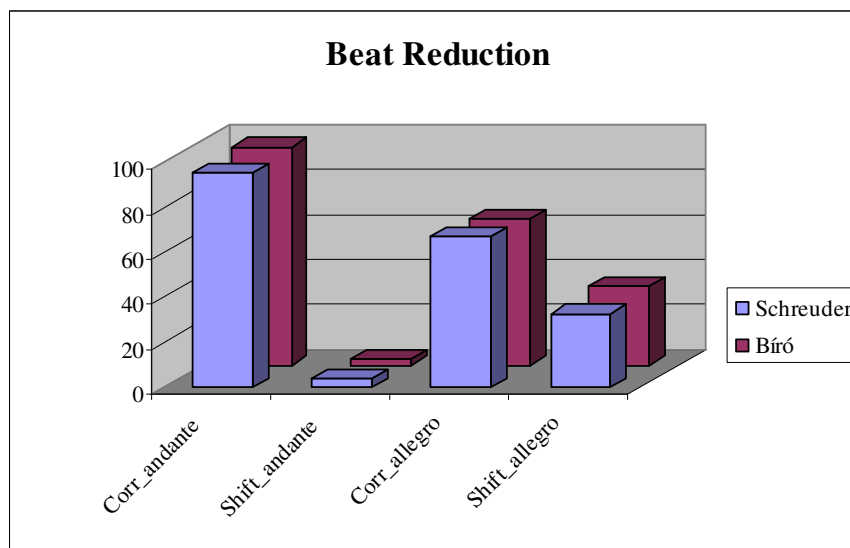
Figure 27 The (horizontal) neighborhood structure of the output candidates (Bíró to appear)



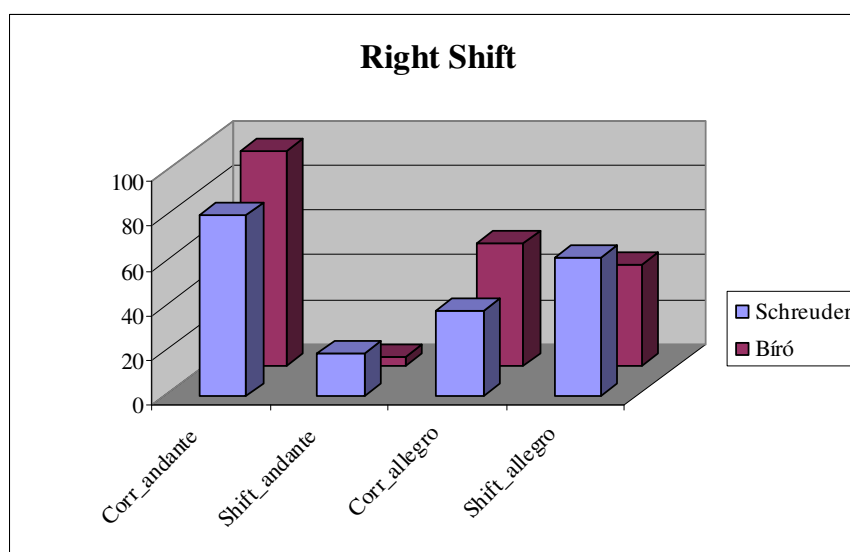
The constraint AFFIX-STRESS demands that a stress-bearing affix (Dutch *-teit*, *-aan*, *-es*, *-in*, *-ist*, etc) must bear main stress in the output. In Figure 28a we see that the simulation of the Beat Reduction type data by B     gives almost the exact frequency distributions of our own speech data. This appears to be a perfect simulation. Figure 28b gives the respective outcomes for the Right Shift data. In the andante cases the simulation is also quite similar to our speech data. The allegro cases, on the other hand, show a difference in direction. However, the outcomes of both studies are centered around the 50%, which means that a different simulation, maybe with some fine-tuning of the rate of the simulation, could give more similar results. For more details, see B     (to appear).

Figure 28 Comparing our data with the Simulated Annealing results in percentages

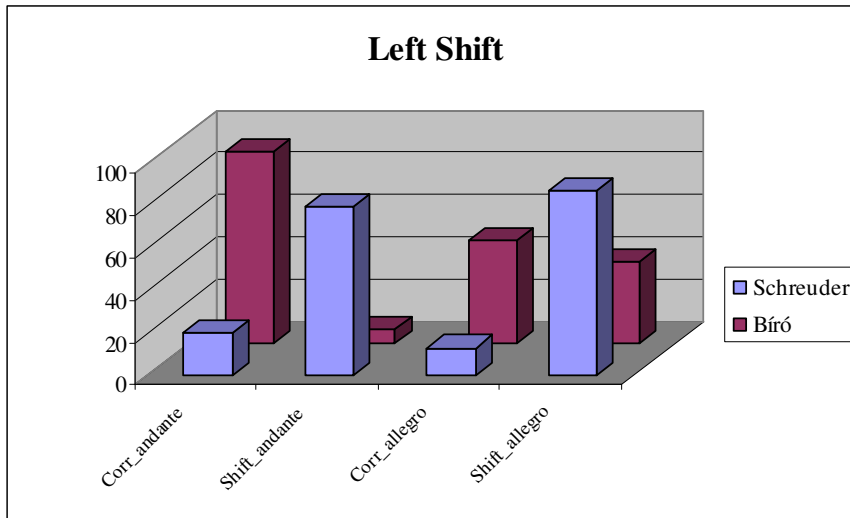
a.



b.



c.



In Figure 28c we see that the simulation does exactly what we expected to find in our own data. Our data, however, do not show what we expected on the basis of the constraint ranking we used, although the shift in frequencies for allegro compared to andante speech is a clear tendency. Here, we can only speculate on the unexpected behavior of the Left Shifts. Maybe the correspondence relation to the base is not active in this type of word, or, as we argued before, the Phrasal Rule interacts with the rhythm patterns here. We are left with the unsatisfactory situation that this question will stay unanswered for now. We cannot blame the simulation, it does exactly what we asked for, and we can therefore conclude that Simulated Annealing is the best OT account thus far for coping with rhythmic variation of secondary stress, because it can deal with frequency distributions of variable rhythm patterns.

The question now is: can we find acoustic evidence for the findings of rhythmic restructuring in fast speech? We will examine this in the next section.

#### 3.5.4. Acoustic analysis

In the current state of phonological research, embodied in e.g. laboratory phonology, much value is set on acoustic evidence for phonological analyses. Studies such as Sluijter (1995) and Sluijter and Van Heuven (1996) provide acoustic correlates for primary stress. According to these studies, duration is the main correlate of primary stress, spectral balance is an important second cue, and pitch also contributes to the perception of stress. Intensity is hardly of any significance. In our study we are concerned with beat reduction and secondary stress shifts and we wonder whether or not the same acoustic correlates hold for secondary stress. Shattuck Hufnagel et al. (1994) and Cooper and Eady (1986) do not find acoustic correlates of rhythmic stress at all. They claim that it is not entirely clear which acoustic correlates are appropriate to measure, since these correlates are dependent on the relative strength of the syllables of an utterance. The absolute values of a single syllable can hardly be compared without reference to their context and the intonation pattern of the complete phrase. Huss (1978) claims that some cases of perceived rhythmic stress shift may be perceptual rather than acoustic in nature. Grabe and Warren (1995) also suggest that stress shifts can only be perceived in rhythmic contexts. In isolation, the prominence patterns are unlikely to be judged reliably. In the remainder of this chapter we will try to find out if we can support one of these lines of reasoning. In other words, are we able to support our perceived rhythmic variability with a phonetic analysis? As a starting point, we adopt Sluijter's claim on primary stress for our analysis of secondary stress. Therefore, we measured all characteristics of main stress, i.e., the duration, pitch (both mean pitch over the whole rhyme, and the maximum pitch in the rhyme), intensity, spectral balance and rhythmic timing of the relevant syllables.

Because Dutch is a quantity-sensitive language, the duration of the relevant syllable rhymes was considered. Onsets do not contribute to the weight of a syllable. In order to make the andante and allegro syllables comparable in duration, the duration was normalized by dividing the durations of the andante words by the durations of the allegro words and then multiplying the durations of the allegro syllable rhymes by this factor. In Table 23a, the mean values of maximum pitch, mean pitch, normalized duration, and

intensity are shown for the syllables which should get secondary stress according to O-O CORRESPONDENCE (hence ‘syllable a’), and in Table 23b for the syllable to which secondary stress can get ‘shifted’ in the restructured rhythm (hence ‘syllable b’). Our measurements would confirm our hypothesis and our auditory analysis, if for syllable a (Table 23a) all phonetic values for ‘No’ in the column ‘Restructured’ were higher than the values for ‘Yes’, and for syllable b (Table 23b) if all values for ‘Yes’ were higher than those for ‘No’. In that case, the subject would realize a word such as *perfectionist* as *perfèctioníst* in andante tempo and as *pèrfectioníst* in allegro tempo. This is not the case. In fact, for syllable a only duration has a higher mean for ‘No’; for syllable b duration is precisely the only correlate with a lower mean for ‘Yes’, so we see exactly the same pattern for both syllables, while these are different syllables, and this is completely unexpected and unexplained.

Table 23a. MANOVA: Descriptive Statistics for syllable ‘a’

Syll_correlate	Restructured	Mean	Std. Deviation	N
a_maxpitch	No	180.171	62.6368	765
	Yes	186.299	62.3552	698
a_meanpitch	No	169.2314	55.99621	765
	Yes	178.0874	58.02636	698
a_duration	No	.1208	.04872	765
	Yes	.0937	.04470	698
a_intensity	No	68.8575	6.14766	765
	Yes	69.8803	6.08864	698



Table 23b. MANOVA: Descriptive Statistics for syllable ‘b’

Syll_correlate	Restructured	Mean	Std. Deviation	N
b_maxpitch	No	177.3691	61.26285	760
	Yes	187.4453	58.10147	703
b_meanpitch	No	167.0725	55.22418	760
	Yes	179.2741	54.93000	703
b_duration	No	.1609	.07698	760
	Yes	.1212	.06039	703
b_intensity	No	66.8939	6.36863	760
	Yes	71.1036	6.11909	703

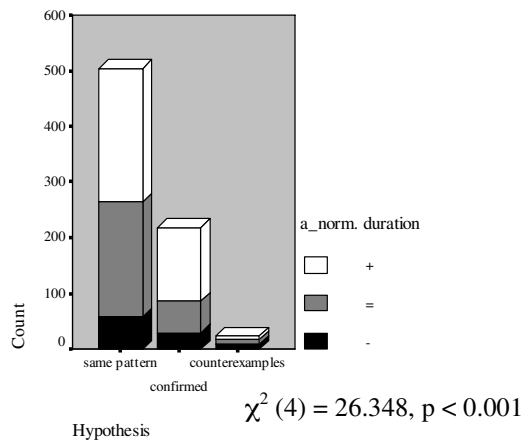
The only possible explanation for the outcomes in Table 23a,b is that the intrinsic values of the segments in the syllables might play a role here. This is because we took all data, andante and allegro, together in this test. Therefore, we separated the measurements of andante and allegro, and we subtracted the acoustic values of the allegro syllables from the values of the same syllables in the andante words. The graphs in Figure 29 give the outcomes for the normalized duration, plotted against the auditory judgments of the word pairs. If syllable a, for instance, has secondary stress in andante tempo and not in allegro tempo, then the outcome is positive, indicated by a ‘+’ in Figure 29a. For syllables a this should correspond to a confirmation of the hypothesis, if the auditory judgments are triggered by these acoustic correlates. We would therefore expect the first bar in Figure 29a to be totally grey (same auditory rhythmic pattern and same acoustic values, or smaller than the just noticeable difference), the middle bar is expected to be white, and the third bar, the counterexamples, should be black. For syllables b in Figure 29b, the same colors apply, yet now white means ‘-’. The white color in both figures indicates the part of the bar in the middle which we expect to be biggest.

The results for normalized duration go in the right direction, and the Chi-Square test gives highly significant differences. In spite of this outcome, to us these results are not convincing. The bars should have had the right color almost totally. For the other acoustic correlates it seems to be random. Therefore, we can conclude that the

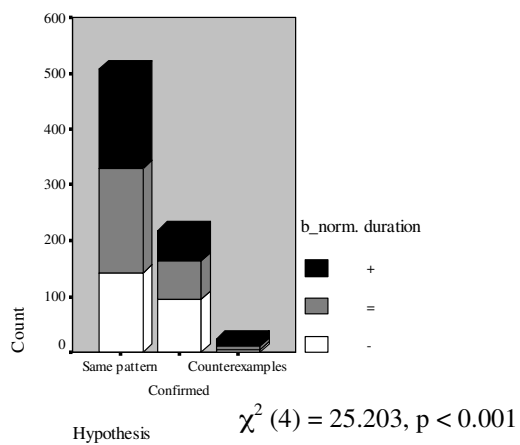
acoustic correlates duration, pitch and intensity are not the relevant correlates of secondary stress.

Figure 29

- a. Chi-Square: Andante values – Allegro values squared with auditory judgements (syllables a)



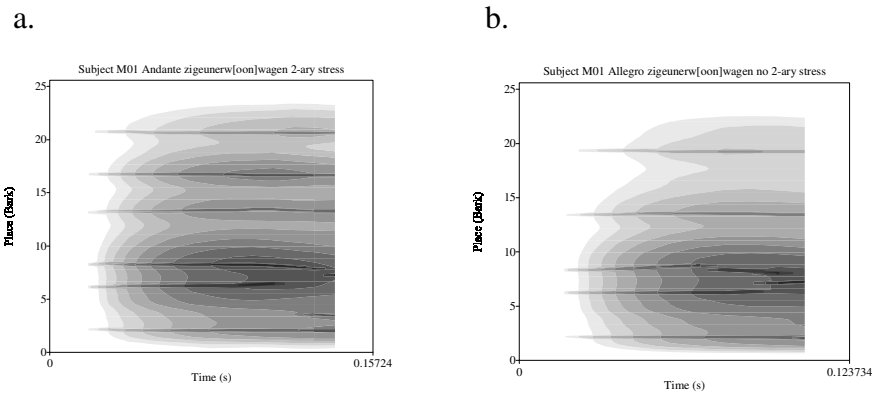
- b. Chi-Square: Andante values – Allegro values squared with auditory judgements (syllables b)



In our pilot experiment, we also considered spectral balance (Schreuder and Gilbers 2004b). Like the other acoustic stress correlates, spectral balance was not the decisive cue for secondary stress. An impressionistic investigation of part of the final data suggests that we cannot expect better results from these data. This impressionistic investigation of spectral balance is described in the following paragraphs.

In order to rule out the influence of the other parameters, we monotonized the data for volume and pitch. Then we selected the relevant vowels and analyzed them as a cochleagram in PRAAT. The cochleagram simulates the way the tympanic membrane functions, in other words the way in which we perceive sounds. In Figure 30 we show two cochleagrams of the vowel [o] in the fourth syllable of, respectively, *zigéunerwòdonwagen* 'gipsy trailer' (Right Shift) in andante tempo and *zigéunerwoonwàgen* in allegro tempo.

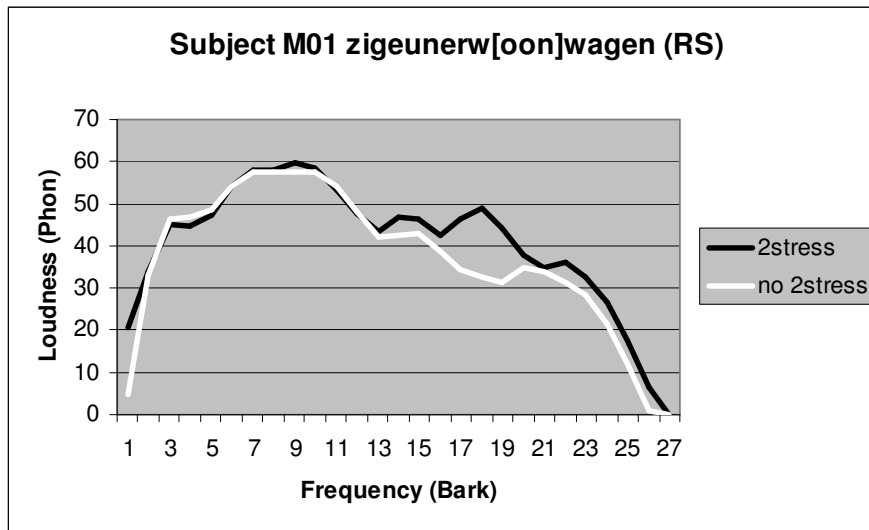
Figure 30 Cochleagrams of [o] in zigeunerw[oon]wagen 'gipsy trailer' (RS)



The cochleagram in Figure 30a (stressed [o]) shows increased perceived loudness in the regions of approximately 5 to 22 Bark in the secondary stressed andante version of [o] in comparison with the cochleagram in Figure 30b (unstressed [o]), indicated by means of shades of gray; the darker the gray the more perceived loudness. This confirms the results of the study of primary stress in Sluijter (1995). If we convert this perceptive, almost logarithmic, Bark scale into its linear counterpart, the Hertz scale, this area correlates with the frequency region of 3 to 10 kHz.

In order to measure perceived secondary stress, we measured the relative loudness in the different frequency regions in Phon.<sup>16</sup> According to Sluijter (1995) stressed vowels have increased loudness above 500 Hz compared to the same vowel in an unstressed position. This can be shown if we take a point in time from both cochleagrams in Figure 30 in which the F1 reaches its highest value (following Sluijter 1995). In Figure 31 the values in Phon are depicted for these points and plotted against the Bark values in 27 steps.

Figure 31 Loudness in Phon



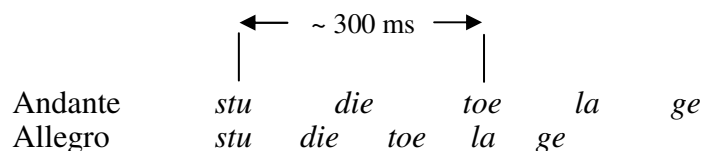
The white line in Figure 31 indicates the pattern of the allegro unstressed [o] in *zigeunerwoonwagen* and the black line indicates the pattern of the andante stressed [o]. We see increased loudness in the region of 13 to 20 Bark, which correlates with the most sensitive region of our ear.

<sup>16</sup> The perceived loudness depends on the frequency of the tone. The Phon entity is defined using the 1kHz tone and the decibel scale. A pure sinus tone at any frequency with 100 Phon is as loud as a pure tone with 100 dB at 1kHz (Rietveld and Van Heuven 1997: 199). We are most sensitive to frequencies around 3kHz. The hearing threshold rapidly rises around the lower and upper frequency limits, which are about 20Hz and 16kHz respectively.

In this item the cochleagrams show the expected differences. It appears to be a mere coincidence, however, because most of the other cochleagrams of word pairs which were perceived as a correspondence pattern in andante tempo and as restructured in allegro tempo were either similar, or different in the opposite direction. The observations do not confirm our auditory analysis and we assume that spectral balance does not characterize secondary stress, as was the case for the other stress correlates.

Therefore, we will look at the data from a radically different perspective: maybe it depends on the listener. We will consider whether the perception of restructuring is based on rhythmic timing. Like music, speech can be divided into a melodic (segment-structural) string and a rhythmic string as partly independent entities. With respect to speech, the melodic string seems to be more flexible than the rhythmic one. Imagine that the rhythm constitutes a kind of metronome pulse with which the melodic content has to be aligned. The listener expects prominent syllables to occur with beats. This behavior is formulated as the *Equal Spacing Constraint*: prominent vowel onsets are attracted to periodically spaced temporal locations (e.g. Couper-Kuhlen 1993, Cummins and Port 1998, Quené and Port 2003). Dependent on speech rate the number of intervening syllables between beats may differ. Suppose the beat interval is constant at 300 ms, there will be more linguistic material in between in allegro speech, e.g. the two syllables *die* and *toe* in *stúdi<sup>o</sup>elàge*, than in andante speech, e.g. only one syllable *die* in *stúdi<sup>o</sup>elàge*. Figure 32 depicts this situation schematically.

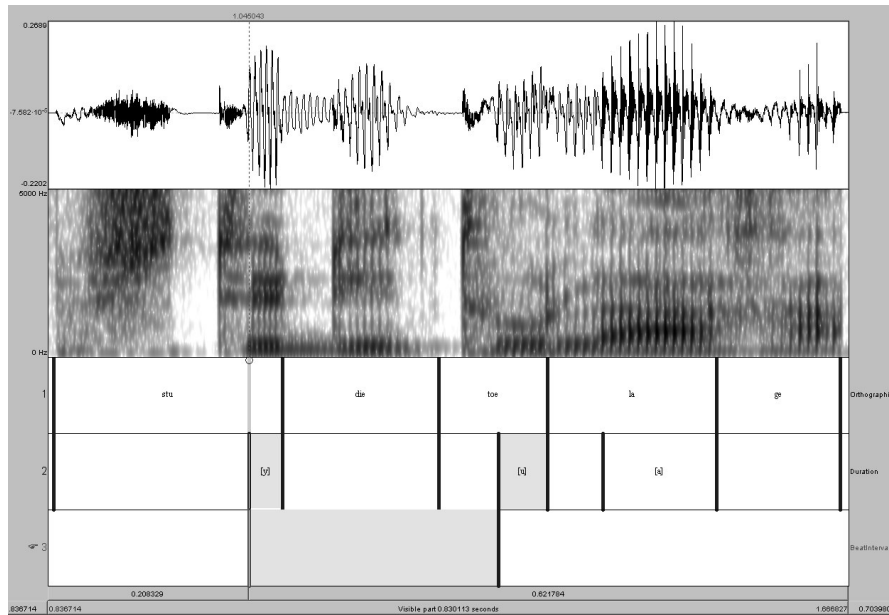
Figure 32 Beat Intervals



In order to clarify the distinction between the duration measurements and the timing measurements, the textgrid in Figure 33 shows the measured intervals for the two dimensions. In this

exemplary textgrid the grey areas in the middle tier give the rhyme durations of the syllables *stu* and *toe*, while the grey area in the bottom tier gives the beat timing interval *between* the vowel onsets of those same two syllable rhymes.

Figure 33 The distinction between the duration and timing measurements

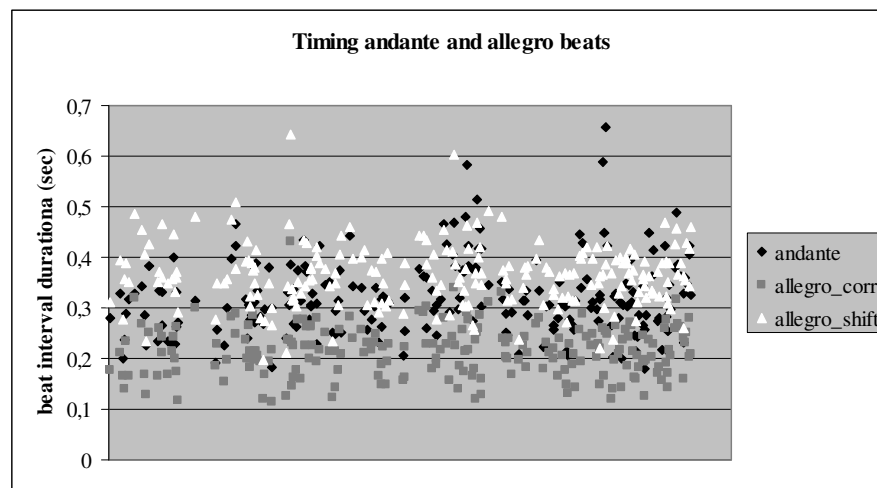


If indeed the perception of secondary stress shifts depends on rhythmic timing, the beat intervals between prominent syllables in *andante* and *allegro* speech are approximately equal. We measured the beat intervals between all possible stress placement sites for all word pairs which were perceived to behave rhythmically the way we predicted in our hypothesis, so the *andante* word conformed to the correspondence pattern, whereas the *allegro* word was perceived as rhythmically restructured.<sup>17</sup> The scatterplot in Figure 34 shows that

<sup>17</sup> Quené provided us with a script that automatically determines the locations of stressed syllables. It examines the energy over the frequency range of the first

the beat interval durations between the Correspondence syllables and the main stress syllables (interval a) in andante tempo, and those of the ‘shift’ syllables and the main stress syllables (interval b) in allegro tempo, are more similar to each other than andante Correspondence intervals to allegro Correspondence intervals. This looks rather promising.

Figure 34 Beat interval durations



However, if we compare this to the boxplots of the same data, in Figure 35, the three groups of beat interval durations appear to be different, even significantly different. Notice that the difference between the ‘andante Correspondence’ interval and the ‘allegro shift’ interval is smaller than the difference between the ‘andante Correspondence’ interval and the ‘allegro Correspondence’ interval. Therefore we examined these differences more closely.

Figure 36 gives the mean values of the three groups of beat interval durations. All three groups differ significantly, as we saw in the boxplots. Nevertheless, the difference scores – the differences of the differences – are also highly significant ( $t(209) = 50.932$ ,  $p <$

---

two formants to identify the sonority rise at the onset of the nuclear vowel. The beat is defined as occurring halfway through the rise, which is similar to the location of the P-center (Morton 1976, Patel et al. 1999).

0.001). This high difference score indicates that listeners do opt for the interval closest to some ideal beat interval.

Figure 35 Boxplots of the beat interval durations

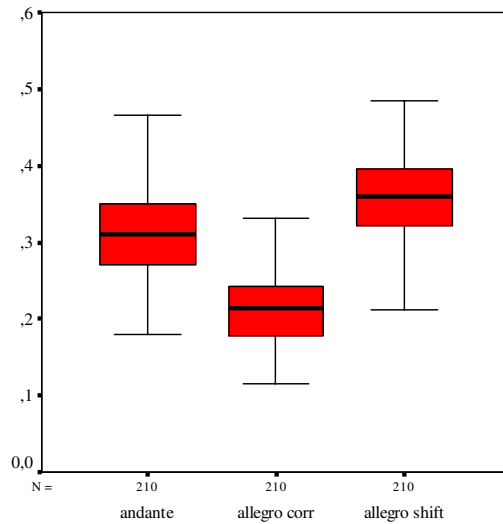
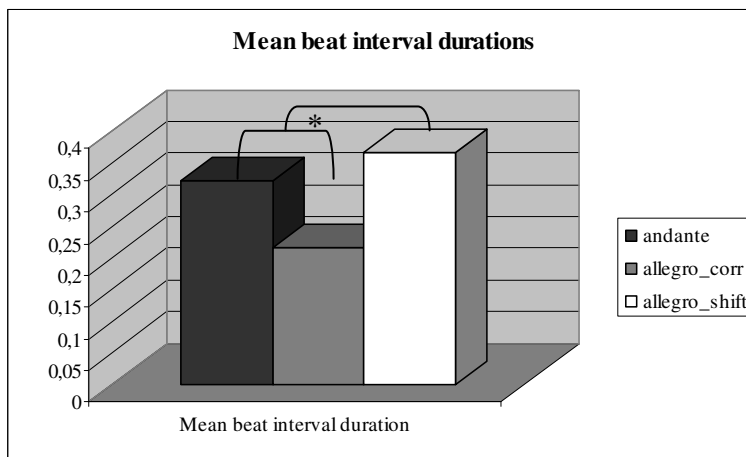


Figure 36 Mean beat interval durations of the three intervals





The fact that the allegro shift intervals do not have exactly the same durations as the andante Correspondence intervals is because speech rhythm is not entirely isochronous (*cf.* Chapter 2), and therefore the intervals between the stressable syllables are not always exactly equal. Just as music can be played in *Tempo Rubato*, which makes the musical melody deviate from metronomic regularity without abandoning the rhythm, so speech rhythm is not a matter of absolute temporal equality (Laver 1994, Fox 2000). Nonetheless, there seems to be an ideal beat interval of between 300 and 400 ms, and that is the interval listeners focus on, at least at the level below the metrical level of main stress. The syllable which is located nearest to this point before the main stress syllable is perceived as rhythmically prominent, and therefore receives a secondary stress in the perception of the listener. In other words, the data supports the idea of a listener-based equivalent of the - speaker-based – ‘Equal Spacing Constraint’, which implies that listeners possess an ‘internal metronome’, which is preset at 200 beats per minute for the secondary stress level, or 100 for main stress rhythm. This is in line with the earlier finding by Couper-Kuhlen (1993:25, note 28) that interstress intervals, i.e. main stress intervals, are typically between 500 and 700 ms in everyday talk. The preference for a meter round 600 ms is remarkable, because in music a beat occurs between about 40 and 300 counts per minute, with a preference for a tempo of around 100 counts per minute, the so-called ‘preferred rate’ – a time interval of 600 ms (Fraisse 1982). This concerns the level of the ‘tactus’, which we take as the musical counterpart of the beat based on linguistic main stress. This implies the same preferred beat intervals for language and music (*cf.* Chapter 2).

The fact that differences in rhythmic structure depend not only on the speaker, but also to a great extent on the listener, is found in music as well. In phonological research, however, this is still a radically new perspective: not all perceived phonological processes have an acoustic realization; we have to consider auditory illusions. As we described in Chapter 2, musical meter is a psychological construct. It cannot be directly measured in a performed rhythm: the listener actively constructs it while listening to music (Honing 2002). Although in music, as opposed to speech, the variable rhythm patterns are in fact mostly measurable, Handel (1993) showed that the same rhythm presented at a different tempo is sometimes

recognized as a different rhythm. Together with the same preferred tempo for speech and music shown above, these findings are an indication that speech and music share some cognitive mechanisms.

### **3.6. Conclusion**

In section 3.3, we presented some different phonological accounts of restructuring within the framework of OT and we tested these accounts with an experiment in section 3.5. Our first conclusion is that phonetic compression cannot be the sole explanation of the different rhythm patterns, because our trained listeners found different rhythm patterns for *andante* and *allegro* tempo.

The results of the Annealing Simulation show the same frequency distributions of rhythm structures as our speech data, except for the Left Shifts, which show unexpected behavior in our data. The model of Simulated Annealing appeared to deal successfully with this kind of variation. Although we will not maintain the hypothesis that there are different grammars, i.e. constraint rankings for different rates of speaking, we have shown that a faster simulation can lead to ‘suboptimal’ outputs as well as optimal outputs, because the simulated ‘speaker’ has less time to search the search space. In their *andante* tempo, data that conform to the correspondence constraints prevail if these are the global optima in the search space, whereas in *allegro* tempo output candidates that obey the markedness constraints can show up more often, as these candidates are local optima. These suboptimal outputs have a more evenly distributed rhythm, and these preferences resemble the preferences of *andante* and *allegro* music. In both disciplines clashes are avoided in *allegro* tempo by means of enlarging the distances between beats.

In section 3.5.4, we attempted to confirm our phonological account with a phonetic analysis. It turned out that none of the phonetic correlates of stress – neither duration, nor pitch, intensity or spectral balance – could identify secondary stress. This is in line with work by Shattuck Hufnagel et al. (1994), Cooper and Eady (1986), Huss (1978) and Grabe and Warren (1995), who all claim that acoustic evidence for secondary stress cannot be found unambiguously.

What we found is that secondary stress is not an acoustic property of speech per se, yet it does exist in the mind of the listener. The listener focuses on time points on intervals of about 300 ms apart, and a secondary stress is perceived on the syllable which is nearest to that point. As opposed to our claims in Schreuder and Gilbers (2004b), the results thus reveal that rhythmic restructuring is more a matter of perception than of production, and is therefore not rhythm, but meter, in the musical sense of the word. The constraint ranking we used seems, in spite of the right predictions it makes for the auditory analysis, to demand a dominant constraint METRONOME for the listener: all stresses are perceived at equally spaced beat locations, with the beat at some preferred rate, which is 200 bpm for secondary stress. From this new perspective for phonologists, we can conclude that it is not always the case that “*meten is weten*”, as we say in Dutch, which means that ‘to measure is to know’ does not always apply.

The reason listeners use this ‘internal metronome’ is probably just a communicative strategy to extract the most important parts from a message. A speaker tries to communicate as much as possible in a short period of time, while a listener tries to select which part of the message is of significance for him. This idea is confirmed by the results of reaction time experiments by Quené (2003), in which he found that subjects’ reaction times were faster if texts were rhythmically regular than if they had an irregular rhythm. For an optimal communication this would mean that if speakers want their audience to pick out the parts they find the most important themselves, conversation partners can best tune their internal beat to each other by speaking in the same tempo.



## Chapter 4

# Recursion in Phonology

### 4.1. Introduction<sup>18</sup>

In this chapter we investigate an instance of phonological recursion; more specifically, we investigate iterative rule application in phonological phrases. We will show that edge-marking processes, such as early pitch accent placement, can be applied recursively to phonological phrases that are embedded in larger phonological phrases. Before addressing prosodic recursion, we will demonstrate that recursion is quite a common phenomenon, not only in linguistics, but in nature, visual art, and music as well.

### 4.2. Recursion

Recursion can be seen as the repeated application of rules in generating a structure, in the sense that it applies to the output of every earlier application. In principle, there is no limit to the extension of the structure. It must be regarded as one of the most powerful mechanisms in generating complex systems, because it is possible to describe the whole system by describing just one layer of the system, since all layers are the same.

Recursion is claimed to be the only uniquely human component of the faculty of language (Hauser, Chomsky, and Fitch 2002). Animal communication systems allegedly lack this rich expressive and open-ended power of human language (Hauser, Chomsky, and Fitch 2002: 1570) that enables us to acquire a complex natural language on the basis of limited data. This statement is rather controversial. In fact Hauser et al. tone down their statement by suggesting that it is

---

<sup>18</sup> This chapter is an extension of Schreuder and Gilbers (2004a). It is supplemented with the acoustic results and an OT analysis.

possible that other animals may develop the same abilities, if recursion in humans evolved from the same cognitive capacity that is also used to solve other computational problems, such as navigation (p.1578). Indeed, examples of recursion can be found everywhere, as will be shown in this section.

Recursion is a very general principle. To begin with, all natural numbers are defined recursively (*cf.* Koster 2003), as shown in Table 24.

Table 24a Recursion in natural numbers

a	$1 = [1]$
b	$2 = [[1] + 1]$
c	$3 = [[[1] + 1] + 1]$
d	etc.

This same mechanism can be shown in the grammars of natural languages. For instance, each sentence can be embedded in a bigger sentence with similar structure, as shown in Table 24b:

Table 24b Recursive sentences

a	[he dreams]
b	[he dreams that [he dreams]]
c	[he dreams that [he dreams that [he dreams]]]
d	etc.

Each sentence can be expanded for ever by making it part of a bigger sentence with the same structure. There is no longest sentence. Mostly, such sentences will not be built of the same words entirely; the structures and types of words must be the same.

#### 4.2.1. Droste effect

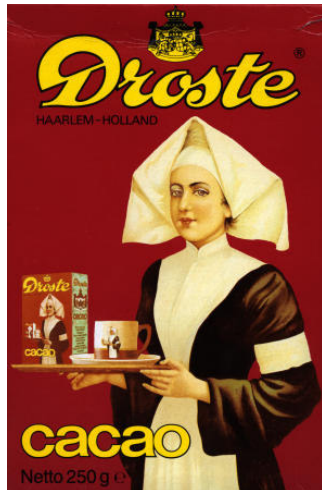
Everyone who uses a computer should be familiar with recursion. For example, the description of folders in a computer can be described by one rule: a folder can contain files as well as other

folders. This simple rule characterizes the recursive system of all folders in the computer.

The visual effect of recursion is shown in the two pictures in Figure 37. Figure 37a depicts the so-called ‘Droste effect’, named after a well-known Dutch cocoa tin. The Droste tin portrays a nurse with a tray on which we see a cup of cocoa and another Droste tin. This Droste tin portrays a nurse with a tray on which we see a cup of cocoa and another Droste tin, etc. Figure 37b depicts another example of visual recursion: the well-known Russian Matruska dolls. If you open the largest doll a smaller doll appears; if you open that doll a still smaller doll appears and so on.<sup>19</sup>

Figure 37 Visual examples of recursion

a Droste effect



b Matruska dolls



<sup>19</sup> In fact Figure 37b visualizes the Matruska dolls as non-recursive, because they are displayed next to each other.

#### 4.2.2. Fractals in nature

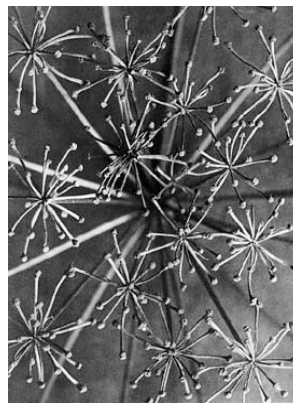
Structures that contain substructures with the same form as the whole, a property known as 'self-similarity', are also found in fractals (Mandelbrot 1977). This kind of recursion is apparent in nature and art. Just think of the leaf of a fern (Figure 38a): on each level the leaf consists of a central nerve with smaller leaves and each smaller leaf is built in exactly the same way. The pictures in Figure 38b by Karl Blossfeldt “Laserwort, part of a fruit umbel” and Figure 38c, a vegetable called *romanesco*, also show examples of recursion in nature. The umbel form and the spirally turreted shape are repeated in their extreme points.

Figure 38 Recursion in nature (see also Kawaguchi 1982, Smith 1984)

a leaf of a fern



b picture Karl Blossfeldt



c *romanesco*





#### 4.2.3. Endless loops in music and visual art

In 1981 the New Wave band The Look released an interesting single called “I am the Beat”. What makes this vinyl single interesting is the fact that it does not end. The final part of the groove is made circular and plays a rhythm that exactly matches the speed of 45 rounds per minute. That is why the final drum part of the song goes on and on, if the single is played on a non-automatic turntable. A similar circular groove was used by The Beatles on their B-side of the album “Sgt. Pepper’s Lonely Hearts Club Band” in 1967. The construction of this record, however, did not contribute to the construction of the song itself. Actually, this is an example of repetitive iteration.

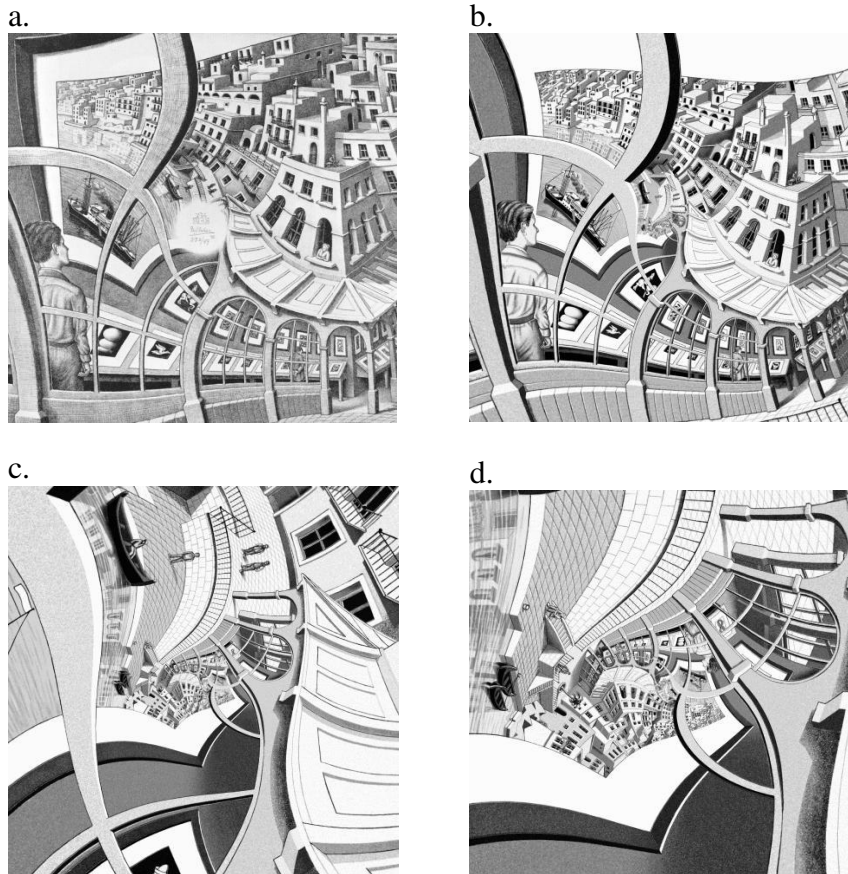
Hofstadter (1979) shows some examples of recursion in music in his book ‘Gödel, Escher, Bach’. An example of a piece that can be extended infinitely, is Bach’s ‘Canon per Tonos’ from ‘Das Musikalische Opfer’, which is also called the ‘endlessly rising canon’, because, from measure 6, it modulates in such a way that it climbs upwards one tone with every repetition. The canon as a whole is in the key C minor, but it concludes - or rather, *seems* to conclude – in D minor. And this “ending” connects smoothly to the beginning again, so one can repeat the whole canon in the key of D minor, which leads to E minor at the end, which again smoothly moves to the beginning. After six such modulations, the original key of C minor has been restored, but one octave higher, and the cycle can start all over again.

An endless loop that includes recursion is demonstrated in Escher’s “Prententoonstelling” (Print Gallery) (1956). In this well-known work a young man is viewing a print in a print gallery. On that print we see a Mediterranean seaport. The print is drawn with a certain elliptic curve, such that one of the buildings in the print, which happens to be a print gallery, at the same time forms the outside of the gallery in which the boy is standing. Escher himself was not able to draw the point where the gallery in the print merges into the big gallery and he therefore drew a white patch in the middle, which contains Escher’s monogram and signature (Figure 39a). In this central spot the recursion of the painting should have appeared.

The mathematicians De Smit and Lenstra (2003) from Leiden University solved the mathematical problem together with their

student Batenburg, and they reconstructed the structure to replace the white patch. Instead of Escher's white patch now a form of the so-called Droste effect appears: on the print in the print gallery we see another gallery with a young man looking at the print on which he sees the gallery with himself looking at... etc (Figure 39b). Thus, the spectator enters an endless loop. Figure 39c and d zoom in on the reconstruction of the white patch, where the recursivity is clearly shown. For animations of the Escher reconstruction, see De Smit and Lenstra's website [escherdroste.math.leidenuniv.nl](http://escherdroste.math.leidenuniv.nl).

Figure 39 Escher's 'Prententoonstelling' (1956) (a.), and its reconstruction (b., c., and d.)



#### 4.2.4. Recursively embedded structures in music

Another branch of art which is known to display recursive structures is music, although the term ‘recursion’ is not very commonly used. Recursion in music has many different manifestations. We find examples similar to fractals, which are referred to by the more general term ‘architectonic music’. In architectonic music large-scale structures echo small-scale aspects, particularly the organization of structures within structures. This is not an exceptional kind of structure in music. Some of these musical structures contain embedded phrases as well (Solomon 1998, see also his website for music examples: <http://music.theory.home.att.net/fracmus.htm>).

Some examples of recursive embedding can be found in Bach’s music. Hofstadter compares Bach’s ‘Kleines Harmonisches Labyrinth’ to a frame story. This piece is written as a labyrinth of quick key changes, in which the listener is soon disoriented. Normally tension builds up in a melody and finally resolves in the tonic, but here the modulations resolve in their own tonic, as if a story within a story ends, while the listener is still waiting for the ultimate resolution in the tonic. This means we hear the harmony recursively: we maintain a mental scheme of the keys, and each new modulation adds a new layer to the scheme. The listener keeps track of the overall key (the tonic) while listening to the local key (the pseudo-tonic) with its pseudo-resolution, and knows when the true tonic is regained. A perfectly recursive structure, center-embedding (*cf.* section 4.2.5), would be if the sequence of keys is retraced in reverse order.

The kinds of recursion in the examples above are not yet real examples of the embedded structures we are looking for. There is another canon of Bach, however, ‘Canon per Augmentationem in Contrario Motu’, canon 4 from the ‘Kunst der Fuge’, in which we find the embedded structure in a part of the melody. The note sequence in measures 30 and 31 contains a melodically identical copy of itself in measure 30, in a twice as fast tempo. The slower melody is in the bass voice, while the faster – embedded – copy is in the leading voice, as can be seen in Figure 40. The only difference is the second E in measure 30 in the leading voice, because a D would have clashed with the C sharp in the bass.

Figure 40 Recursion in Bach's *Canon per Augmentationem in Contrario Motu*

Whole phrases can be recursively embedded. Koch (1983) and Rothstein (1989) describe the phenomenon of phrase expansion, which is defined as the transformation of a phrase into a longer phrase, by adding more notes. These transformations are perceived as different representations of the same phrase. This phrase is the structural skeleton making up both phrases. In the experience of the listener a phrase expansion departs, often quite unexpectedly, from a fixed point of reference, and returns to it after a detour, bringing resolution and reassurance. The original phrase usually has a very regular hypermeter, while an expansion temporarily suspends this hypermeter without actually breaking it. Listeners can often “hear through” the expansion to the underlying hypermeter.

Not all phrase expansions are recursive in the sense we defined above. Rothstein (1989) distinguishes two kinds of phrase expansion that have recursive manifestations: external phrase expansions and internal phrase expansions.<sup>20</sup> External phrase expansions are an addition of subordinate material either before or after the basic phrase, leaving the basic phrase more or less unaffected. Internal expansions, however, add length within the basic phrase itself and are often literal or varied repetitions within the phrase. “Small” prefixes and “small” suffixes are external phrase expansions. A small prefix is less than a phrase; it is an incomplete phrase, an accompanying figure that sets the stage for a melodic entrance (Reicha 1814). This can also be found in vocal music, where often the melody is introduced by means of a short stretch of melody in the accompaniment just before a solo entrance, as a kind of “pre-imitation”. Measures 9 to 17 of the first movement of Schubert’s

<sup>20</sup> Rothstein does not mention the term ‘recursion’. The way he describes the phenomenon, however, is very similar to how recursion is defined above.

“Unfinished” Symphony, No. 8 in B minor, give an example of an expanded phrase with a small prefix (see Figure 41). The phrase begins with a prefix, which does not have a cadence of its own but instead moves to the accompaniment for the melody of the main phrase. The main phrase starts at measure 13. We can compare this kind of phrase to a recursive noun phrase in language, with two comparable adjectives: [*international* [*diplomatic organizations*]<sub>NP</sub>]<sub>NP</sub>; the phrase has two starting points, but only one end, which means that it constitutes a recursive phrase.

Figure 41 Recursive phrase: Schubert’s “Unfinished” Symphony, No. 8 in B minor, first movement, mm. 9-17

The musical score illustrates a recursive phrase structure in Schubert's "Unfinished" Symphony, No. 8 in B minor, first movement, measures 9-17. The score is in 3/4 time and B minor. It is divided into three systems. The first system (measures 9-11) is labeled 'a) PREFIX' and 'm. 9 Violins'. The second system (measures 12-14) is labeled 'Ob. Cl. PHRASE:'. The third system (measures 15-17) is labeled 'etc.'.

Like prefixes, phrases can be expanded with suffixes. Suffixes are not forward-moving, developing structures; rather they form a stasis, an extension of a goal already reached (Riemann 1902). Small suffixes after a full cadence function as an expanded repose or codetta. A very common example is the expansion of a (full) cadence with another (full) cadence. The end of the suffix can serve as the

end of the basic phrase. The phrase has just one beginning and two or more endings, and can therefore be seen as a recursive, embedded phrase, as shown in Figure 42.



The phrase can be embedded in several phrases, because suffixes themselves may have suffixes. Simple examples are found in musical phrases in which the final cadence is repeated. In these simple kinds of recursion one speaks of ‘tail-recursion’, because only the last part of the structure is recursive. In “The wind cries Mary” by The Jimi Hendrix Experience (1967) the movement towards the tonic F is repeated six times: E flat – C/E – F.

Internal phrase expansions, as said above, add extra notes within the phrase itself. A melodic line in this kind of expansion may either repeat part of the basic phrase or may deviate from it. Often a part of the bass line is actually repeated. An internal phrase expansion can be recognized if the total length of the phrase exceeds the phrase lengths in the rest of the piece and if other phrases of normal length do not contain similar repetitions. Examples can be found in phrases with a deceptive cadence (V-VI) followed by a delayed full cadence. According to Rothstein (1989), this should not be seen as a suffix, but as an internal expansion, because the true end of the phrase arrives only with the authentic cadence (V-I). Deceptive cadences indicate expansion if they occur where a full cadence is expected, if the full cadence actually follows, and if this full cadence may be satisfactorily substituted for the deceptive one. Figure 43 gives an analysis of an internal phrase expansion in Haydn’s Quartet in D minor, Opus 76, No. 2, second movement, measures 47-51; 61-62. It is easily recognized as an expansion, because the same phrase has occurred earlier in the movement (measures 36-39, not shown) with an authentic cadence in its fourth bar.

Figure 43 Haydn's Quartet in D minor, Op. 76, No. 2, 2<sup>nd</sup> movement, mm. 47-51; 61-62 (Rothstein 1989)

47

48

50

61

b)

pizz.

arco

V 6 4 - 7 3

V 6 4 - 7 3

I

(deceptive cadence—long expansion leads to authentic cadence shown in b)

#### 4.2.5. Computing recursion

All examples mentioned are instances of regular context-free recursion, which is easily recognized by Finite State Automata (FSA). A different kind of recursion, center-embedding, is impossible to describe for FSAs, as the relationships between the embedding constituents on either side of the embedded constituent represent an unbounded quantity of information, which cannot be represented by finite states (*cf.* Chomsky 1959, Nederhof 2000). Center-embedding is found in human syntactic parsing, in sentences like *the rat [that the cat [that the dog chased] killed] ate the cheese* (Chomsky and Miller 1963). In theory people could pronounce such constructions, however, in practice people cannot store that much information in their working memory effectively till the verb of the sentence is parsed. We have not found evidence of such constructions in prosody thus far, whereas we could think of center-embedded musical structures (*cf.* section 4.2.4).

In the remainder of this chapter we will investigate whether the regular context-free kind of recursion can be found in prosody.

### 4.3. Recursion in phonology

#### 4.3.1. Strict Layering and recursion

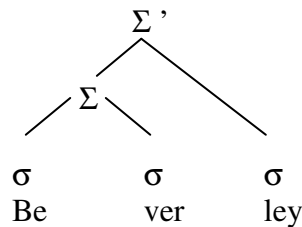
In the previous section we have shown many examples of recursion in nature, visual art, and music. Recursion has been called the ‘only uniquely human component of the faculty of language’ (Hauser et al. 2002). Hauser et al. make this statement in reference to syntax. Recursion refers to rules which are capable of repeated application in generating a sentence. In principle the number of prepositional phrases that may occur after a noun in a noun phrase is unlimited: *the American in the desert on a horse with no name*, in which *with no name* is a PP embedded in the PP *on a horse*, and you can always add a sentence to a sentence within a sentence as exemplified in the present sentence, and in Table 24b. As Crystal (1991, p.292) puts it, the application of recursive rules is the main formal means of



accounting for the creativity of language: by using this device, an infinite set of sentences can be generated from a finite set of rules.

In phonology, things seem to be different. Although iterative rule application is proposed for e.g. foot assignment, prosodic building rules seem to be limited in that sense. One cannot freely add e.g. onsets or nuclei to a syllable or syllables to a prosodic word. One of the rare occurrences of the incorporation of a prosodic domain within the same prosodic domain can be found in Selkirk (1980, 1984), who proposes a foot within a foot, constituting a super-foot, in order to account for dactylic patterns in rhythmic structures, as depicted in Figure 44.

Figure 44 Recursive foot (Selkirk 1980; 1984)

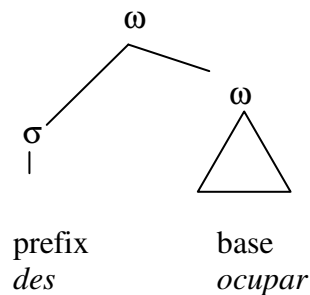


Selkirk's representation seems to give an argument for recursivity in phonological domains, although she circumvents recursion by calling the maximal foot a super-foot, indicating that both feet are of different categories. In her approach this kind of recursivity of phonological domains is restricted to the dactylic foot. Her approach is challenged by the fact that there are alternative representations without the need for recursivity. Dresher and Lahiri (1991) propose a ternary branching tree, and Kager (1994) maintains binarity and leaves the third syllable unparsed.

The limitations to the prosodic hierarchy are reflected in the Strict Layer Hypothesis (Selkirk 1984), of which one of the fundamental assumptions is that prosodic structure is not recursive. A mismatch thus exists between syntactically recursive constituent structure and the linearly segmented structure in prosody.

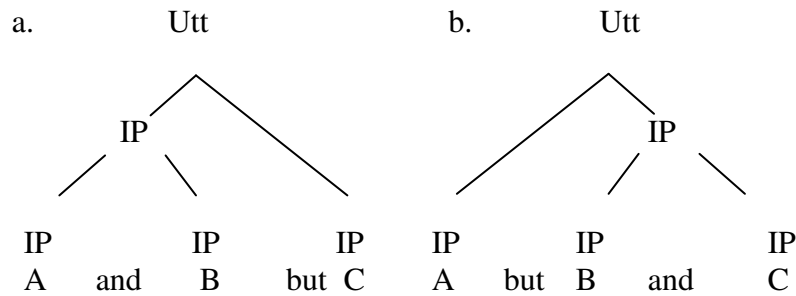
The Strict Layer Hypothesis has been criticized. Several phonologists, such as Itô and Mester (1992), have shown that in many cases it is violable; the assumption of non-recursivity has been challenged by a number of authors, in particular with regard to prosodic words with affixal clitics (Zec and Inkelas 1991, Booij 1996, Vigário 1999 among others). Vigário, for example, represents prefixed words in European Portuguese as a syllable and a prosodic word within a prosodic word, as depicted in Figure 45. Notice that similar prosodic domains should display similar phonological behavior. In European Portuguese word-internal vowels can be reduced, word-initial vowels cannot. Although the <o> in *desocupar* 'disoccupy' seems to be a word-internal vowel, it actually fails to undergo vowel reduction, because it is still the initial vowel of a prosodic word. This cannot be explained if recursion is not allowed in the prosodic hierarchy.

Figure 45 Recursive prosodic word (Vigário 1999)



Examples of recursion in larger prosodic domains like the Intonational Phrase can be found in Ladd (1986, 1996), and references therein. The Intonational Phrase (IP) is the domain of a perceptually coherent intonational contour (Shattuck-Hufnagel and Turk 1996). These Intonational Phrases are dominated by the Utterance (Utt), which is often described as the phonological counterpart of the syntactic sentence. It is described as the largest span of application of phonological rules (Selkirk 1978, 1980, Nespor and Vogel 1986, Hayes 1989).

Figure 46 Recursive intonational phrase (Ladd 1986, 1996)



Ladd represents different kinds of conjunction as in Figure 46. Phonetic differences, such as pause-duration cues and declination slope reset motivate the division into three Intonational Phrases, yet the *but*-boundaries are significantly stronger than the *and*-boundaries. The difference in strength between the conjunction of (*A and B*) as opposed to (*(A and B) but C*) is visualized in Figure 46 by embedding (*A and B*) in a dominating IP. Without recursion, we cannot explain the phonetic differences between the two kinds of conjunction. The following sentence, which has the structure of Figure 46a, exemplifies these different conjunctions: *Aretha Franklin is a soulful singer, and Carole King is an excellent composer, but Florence Foster Jenkins amuses us the most.*

Similar arguments to the ones above can be found in Ladd (1992), Inkelas (1989) and McCarthy and Prince (1993a,b). These arguments led Selkirk (1995b) to replace the Strict Layer Hypothesis with a series of four separate constraints, one of which is Non-Recursivity: No  $C_i$  dominates  $C_j$ ,  $j = i$ . This is a violable constraint, in terms of standard OT (Prince and Smolensky 1993).

Since recursion is very common in syntax, the primary source of evidence for instances of recursion in phonology is probably provided by phonological rules that operate over (morpho-) syntactically defined recursive domains. Notably, in the phonological domains where morphosyntax does not play a role, no

evidence has been found for recursive structure - like the syllable domain - or the arguments are weak, like for the domain of the foot. The phonological phrase is one instance of a domain of which the phrase breaks typically coincide with the edges of morphosyntactic phrases (Selkirk 1984, Nespor and Vogel 1986). Although there is no consensus on what exactly constitutes the phonological phrase, we follow Selkirk, who assumes that the phonological phrase aligns with either the left or the right edge of the head of a maximal projection which is not lexically governed, i.e. it groups a phrasal head together with its adjacent modifiers and functional elements (Selkirk 1995b).

#### 4.3.2. Research question

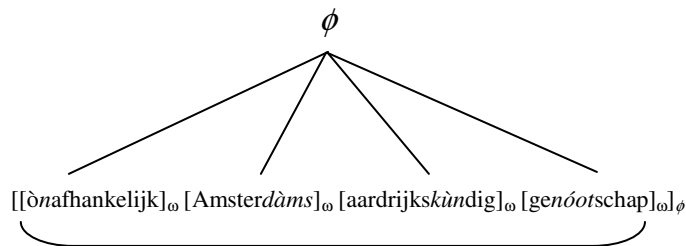
In this chapter we will show that recursion can be found at the higher domains in the prosodic hierarchy. To investigate this, we examined Dutch phonological phrases, consisting of either a noun modified by one adjective, of the type *aardrijkskundig genóotschap* 'geographical society', or by two adjectives, of the type *Amsterdàms aardrijkskundig genóotschap*, i.e. a recursive noun phrase (the accents on the adjectives indicate the main stress position in citation form.). Syntactically, this kind of phrase can in principle be infinitely extended with adjectives: [*onafhankelijk* [*Amsterdàms* [*aardrijkskundig genótschap*]<sub>NP</sub>]<sub>NP</sub>]<sub>NP</sub> 'independent Amsterdam geographical society'.

The first type of phonological phrase, with one adjective, is known to display early pitch accent placement (Shattuck-Hufnagel 2000, Shattuck-Hufnagel et al. 1994) as a means of signaling a phrasal boundary to give the listener a cue to the prosodic structure of the spoken utterance. This phenomenon was first referred to as Iambic Reversal (Liberman and Prince 1977) and it is also known as stress shift, the (English) Rhythm Rule (Liberman and Prince 1977), or the Phrasal Rule (Hayes 1984, Shattuck-Hufnagel 2000). In contradistinction to these stress movement accounts, Gussenhoven (1991) proposed that the phenomenon is not movement of lexical main stress, but a combination of two events: the occurrence of a phrase-level intonational prominence on the earlier full-vowel

syllable, and the non-occurrence of a pitch accent on the later main-stress syllable. Horne (1990), Grabe and Warren (1995), Shattuck-Hufnagel et al. (1994), and Vogel, Bunnell, and Hoskins (1995) subsequently showed that that indeed was the case. We can therefore say that Hayes' Phrasal Rule is a boundary marking phenomenon, which marks phrasal boundaries. As we will show, however, the boundary marking not only depends on pitch, but also on duration and timing (*cf.* section 4.4.5.2). Moreover, Quené and Port (2003) found that rhythmic timing also has a strong influence on the process.

The question now is what kind of prosodic structure has to be assumed for the second type of phrase, modified by two adjectives, the syntactically recursive noun phrase. If the non-recursivity assumption holds and these phrases are non-recursive, then they must have a flat, linear structure, and no early accent will occur on the intermediate adjectives: [*ònafhankelijk Amsterdàms aardrijkskùndig genóotschap*], as depicted in Figure 47.

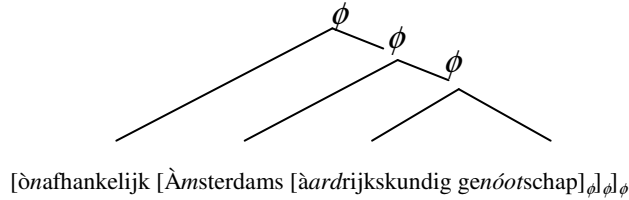
Figure 47 Non-recursive phonological phrase



If, on the other hand, an early accent can be realized on these intermediate adjectives, we have to assume a nested, recursive phrase structure, indicating two or more left boundaries within the same phonological phrase, i.e. a nested prosodic structure: [*ònafhankelijk*

[*Àmsterdams* [*àardrijkskundig genóotschap*]]], as shown in Figure 48.<sup>21, 22</sup>

Figure 48 Recursive phonological phrase



To investigate whether these kinds of phonological phrases can indeed be produced with a recursive prosodic structure, with the Phrasal Rule applying two or more times, we conducted an experiment, which is described in the next section.

#### 4.4. The experiment

##### 4.4.1. Task design

In order to get as close to spontaneous speech as possible, we used the Map Task (Brown et al. 1984) to build our corpus in a controlled way. The Map Task is originally a cooperative task involving two participants, used to build dialogue corpora. We adapted the original design somewhat to our own requirements. The subject and the experimenter sat opposite one another, the subject sat in the soundproof studio behind a glass window, and each had a map which the other could not see. The subject had a map consisting of a starting

<sup>21</sup> Selkirk (1995a) herself gives some examples of a similar kind in English: [[nòrthern][Càlifornia wínes]] as opposed to the right-branching phrase [[nòrthern Califòrnia] wínes], but without going into its recursivity. We hypothesize that these syntactically recursive noun phrases can be realized as recursive phrases in prosody as well.

<sup>22</sup> In terms of computational linguistics, this is a regular context-free right-recursion, which is quite easy to describe for finite-state automata (*cf.* Chomsky 1959, Nederhof 2000).

point, an endpoint and some landmarks, labeled with their names, on the route. The phrases of interest were two of the landmarks, the rest were fillers (see Figure 49). The experimenter's map only had the starting point drawn on it, which makes the experiment more or less 'double blind'.

We made fifty different maps, with two landmarks of interest on each map, which makes a hundred phrases in total. Each map had four fillers and the phrases of interest never appeared as a starting point or endpoint. Each map contained one syntactically recursive phrase landmark [Adj [Adj Noun]] and one non-recursive, non-corresponding phrase landmark [Adj Noun]. The subjects never saw two corresponding phrases.

#### 4.4.2. Subjects

We tested 24 subjects, ten men and fourteen women, aged 19 to 28. Most of them were law students, with Dutch as their mother tongue. Ten subjects were brought up in the northern provinces of the Netherlands, nine of them came from the center, three from the west and two from the south. One subject had grown up in the Netherlands Antilles, and Dutch was not her mother tongue, though she learned it in her childhood. We found no differences in the characteristics of interest, so we kept her in the experiment. We did not find any regional influences on the results either. The subjects were unaware that it was a linguistic experiment.

#### 4.4.3. Method

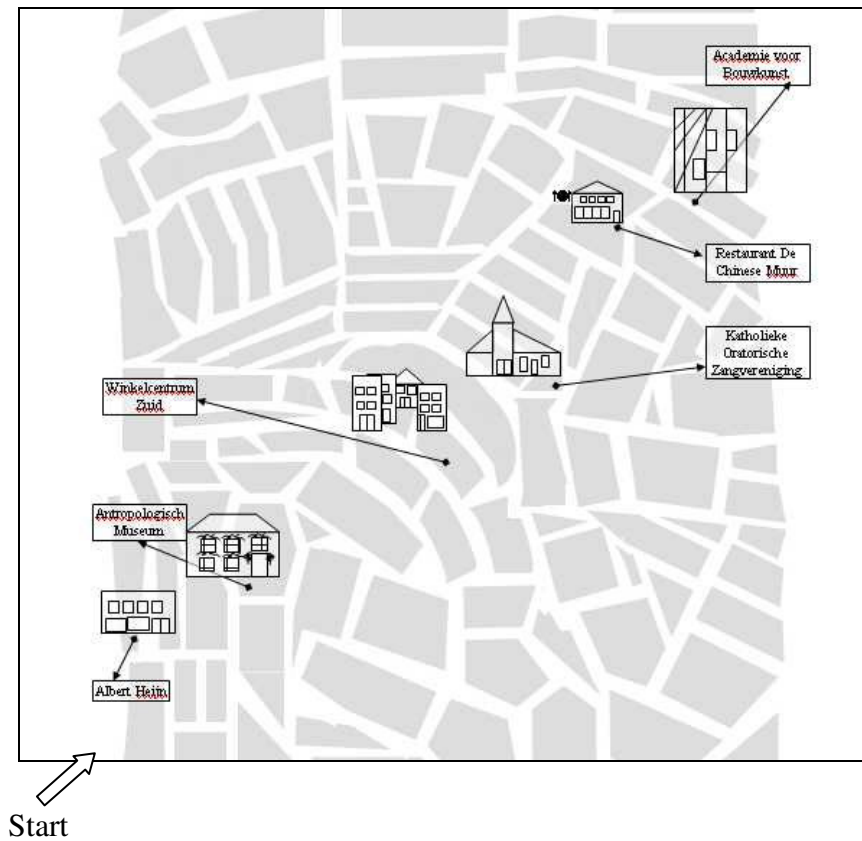
The subjects were told that their goal was to lead the experimenter from point A to point B on the map, leading past all landmarks on the map, and they were supposed to mention all the landmarks they came across. The picture in Figure 49 shows part of such a map. The following text is a translated fragment of a map description of the map in Figure 49 by the male Subject M04. (Bold phrases are the English glosses of the landmarks depicted on the map.)

Question: *'We are at **Albert Heijn's** and I want to go to the **Academy of Architecture**. Can you tell me the way?'*

Answer: *'Ehmm, - let's see - in front of **Albert Heijn's** – you turn left, – and then you take eeh – then go straight ahead and then you turn eeh – second street left, – annnd, then just walk straight ahead, then you come across the **Anthropological Museum**, [...] – eehm – when you stand in front of the **Anthropological Museum** you just walk straight on, – annnd eehmm – then you come across eeh first a T-junction then you keep on walking straight ahead, at the crossroad you just cross it, – eehm – then you again arrive at a T-junction – kind of a T-junction – then turn right, – annnd eehmmm – [...] then you follow a slight curve to the right, eehmm, [...] after that you turn to the first on your left immediately, then you arrive at the **Shopping Centre South**, eehhm – then immediately to the right again, annnd – tooooo the left, [...] ehm you walk straight ahead, then you come across the church of the **Catholic Oratory Choral Society**, eehm arrived there ehm you turn right, annnd after that immediately to the left and then you walk up that eeh that street that you were walking already – [...], and then you arrive at the **Academy for Architecture** and before that you came across **Restaurant the Chinese Wall**.'*



Figure 49 A map used for the map-task



The experimenter did not interfere. Afterwards the subjects were asked to read the adjectives aloud in citation form, within the sentence *Ik spreek nu het woord ... uit* 'I now pronounce the word ...'.

All data were recorded with a Sennheiser MKH 40 Microphone (mono), on a Sony DTC-57ES DAT-recorder, with Fuji Digital Audio Tapes. The sound files were digitalized with Cool Edit Pro at a 22050 Hz sample rate, mono with 16-bit Resolution, normalized to 100%, and saved as .wav files (Windows PCM). The phrases of interest were extracted from the sound materials; the same procedure was used for the citation form words.

For the auditory analysis five trained listeners judged the data auditorily and indicated on which syllables in the adjectives they perceived word accent. They were free to indicate more than one accent per adjective, which meant that words could be double pitch accented. A majority judgment of the five trained listeners was decisive; it turned out that there was consensus among three listeners on most data. We ran Chi-Square tests on the statistics.

For the acoustic analysis we analyzed the data in PRAAT (Boersma and Weenink 1992-2006). We measured fundamental frequency in Hz (maximum and mean), duration in seconds and intensity in dB of the rhymes of the two syllables of interest, for each phrasal adjective. Furthermore, we measured spectral balance in a small set of the data and we measured rhythmic timing between the perceived accents in the phrases. The same measurements were made on the words in citation form, and we compared the values of the words in phrases and in citation form with T-tests. A Multivariate Analysis of Variance showed us which acoustic cues were responsible for the perception of a pitch accent, and moreover, we did Chi-square tests to compare the values of the words in phrases and in citation form in relation to the perceived accents. Inter-accent intervals were compared with T-tests as well.

#### 4.4.4. Data

As pointed out above, the data consisted of one hundred phonological phrases, half of which were [Adjective Noun] combinations and the other half corresponding [Adjective [Adjective Noun]] combinations. Table 25 shows a selection of our data.<sup>23</sup>

---

<sup>23</sup> Some examples from the experiment can be downloaded as MP3-files from <http://home.planet.nl/~schre537/sounds.htm> or [www.maartjeschreuder.nl](http://www.maartjeschreuder.nl).

Table 25 Data

<i>Aardrijkskundig genootschap</i>	'geographical society'
<i>Amsterdams aardrijkskundig genootschap</i>	'Amsterdam -'
<i>Diplomatieke organisaties</i>	'diplomatic organizations'
<i>Internationale diplomatieke organisaties</i>	'international -'
<i>Regionale dagbladers</i>	'regional daily press'
<i>Algemene regionale dagbladers</i>	'general -'
<i>Socialistische partij</i>	'socialist party'
<i>Progressieve socialistische partij</i>	'progressive -'
<i>Psychiatrisch ziekenhuis</i>	'psychiatric hospital'
<i>Academisch psychiatrisch ziekenhuis</i>	'university -'

In order to minimize the influence of pure regular rhythm instead of prosodic structure, for example eurhythmicity effects of the Quadrisyllabic Rule (Hayes 1984), we varied the number of syllables between the accentable positions in the words from 1 to 7.<sup>24</sup> We also avoided stress clash effects, and contrast effects of phrases ending in a similar suffix or contrasting phrases on one map.

The phrases we used are of a type which can undergo so-called Rhetorical Retraction (Gussenhoven 1983, 1984). Rhetorical Retraction in fact refers to the same phenomenon as early pitch accent placement, but Van Bezooijen (2001) shows that speakers use Rhetorical Retraction mostly as a propagandistic speech style. Gussenhoven (1983, 1984) shows that the effect of style is significant but small, in the least rhetorical style. In our experiment we made the speech context as neutral as possible, in order to show that early accent placement is not only a stylistic phenomenon, but also a structural device. With the map task we were guaranteed a non-commercial, neutral context, and we will not use the term Rhetorical Retraction for our data.

<sup>24</sup> Recall that we counted the Phrasal Rule, which is one of Hayes' original eurhythmicity rules, as a boundary marking principle, no longer as a (eu)rhythmicity rule.

Ten out of the fifty recursive NPs can be interpreted as if the first adjective modifies the second, instead of the noun, as possible in e.g. *progressief individualistisch verbond* ‘progressive individualistic union’. If subjects indeed interpreted a phrase as ‘progressively individualistic union’, the recursive prosody was not expected to be realized in that phrase. We will look into the data whether phrases with such an ambiguous interpretation behave differently from the rest.

We divided the subjects over five different map sets, so each subject read ten maps, which means ten recursive and ten non-recursive phrases for each subject. This resulted in about 550 spoken phrases in total. An impressionistic observation reveals that the subjects mostly pronounced the names of the landmarks in focus, and most of the times it was before a comma with a pause, pronounced with a so-called continuation rise (L-H%) or before a full stop, with the so-called declarative contour (L-L%) (Pierrehumbert 1980, Gussenhoven et al. 1999). Some subjects occasionally repeated the phrases. The repetition was then most of the time out of focus and sometimes with a different rhythmic pattern. Others unfortunately missed some of the phrases. In the results section we only report on single, first uttered, utterances of a phrase for each subject.

#### 4.4.5. Results

##### 4.4.5.1. Auditory results

Five trained listeners indicated where they perceived pitch accents on the adjective in the non-recursive noun phrases and on each of the two adjectives in the recursive noun phrases. When a majority indicated they perceived a pitch accent on a certain syllable, this syllable was appointed a 1, the other potential pitch accent site was assigned a 0. In most of the cases the listeners agreed on the pitch accent position as either 0 or 1. Figure 50a shows the percentages of perceived pitch accents perceived on the early syllable, on the main stress position, and also the percentage of the cases that were judged to be double accented. Figure 50b depicts the results for the words in

citation form, where the main stress position is accented almost always.

The graph in Figure 50a clearly shows that, although there is a strong preference for the subjects only to accent the main stress syllable, early accent placement is also a strong tendency. For the non-recursive phrases almost 30% displayed an early accent. This is the same percentage Van Bezooijen (2001) reports for rhetorical expressions, while we had a totally neutral context. This already is a surprising result.

The most interesting result for this study, however, is the early accent bar of adjective 2 of the recursive phrases in Figure 50a. Although these adjectives were not the initial words of the longer phonological phrases, they still received an early accent in 22% of the phrases. This result seems to confirm our hypothesis that these syntactically recursive phrases can also be recursive prosodically (Schreuder and Gilbers 2004a).

Subjects and items differ greatly in their behavior and patterns, however. The standard deviations are very large or even maximal for the items. This means that some items never conformed to the Phrasal Rule, and others, on the other hand, always did. We could not find any systematic characteristic in the items which did or did not show many shifts. It seems not to be a rhythmic phenomenon: the number of syllables between accentable syllables in a phrase does not influence the number of early pitch accents on that phrase. Gussenhoven (1983), on the other hand, found that if one of the interstress intervals contains a low number of syllables, the propensity for early accent placement increases.

Another interesting finding is the relatively high percentages of double pitch accents on phrasal adjectives, i.e. a pitch accent both on the early syllable and the main stress position of the word. Shattuck-Hufnagel (2000) reports a similar finding. These two findings are an argument for the view that it is not a matter of stress shift, but a phrasing phenomenon, which involves accents. This means that the eurhythmy rules (Hayes 1984) do not apply to these data, at least not when we base rhythm on syllable counting.

As said in section 4.4.4, some of the recursive NPs have a possible interpretation in which Adjective 1 modifies Adjective 2. In

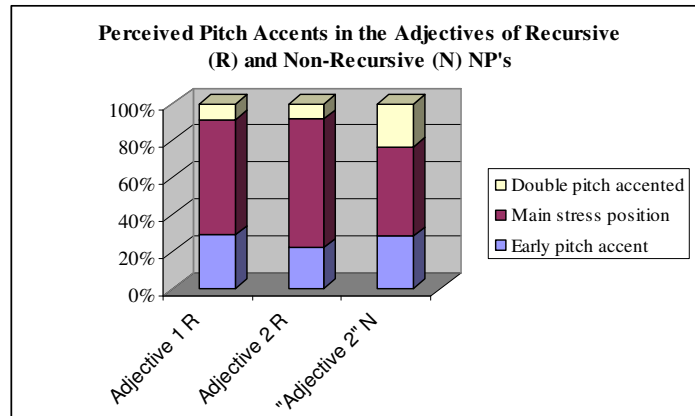
the data we see quite a large amount of non-restructured, and non-recursive realizations with an early accent only on Adjective 1, but also a number of recursively early pitch accented ones. This means the observations for these phrases do not differ much from the overall pattern.

Besides the subject and item dependencies, Pearson Chi-Square tests show that the proportions of early accents and main stress positions in the phrase and in citation form are significantly different ( $\chi^2$  (df 2) = 117.209,  $p < 0.001$  for the non-recursive NPs,  $\chi^2$  (df 2) = 54.552,  $p < 0.001$  for the recursive NPs). The adjectives in citation form had a nearly 100% score of pitch accents in main stress position. This means there is a substantial proportion of the data in which an early accent is perceived on adjective 2, both for the non-recursive and the recursive phrases. However, the difference between the proportions of corresponding adjectives of the non-recursive and recursive phrases is also highly significant ( $\chi^2$  (df 2) = 20.393,  $p < 0.001$ ). Note that the Chi-Square value of the non-recursive vs. recursive test is much smaller than the values of the phrase vs. citation form tests, which indicates that the difference between the patterns of the corresponding adjectives in phrase-initial and phrase-second position are much smaller than the differences between the adjectives spoken in phrase-second position and in citation form. Nonetheless, we can conclude that listeners do perceive prosodic recursion.

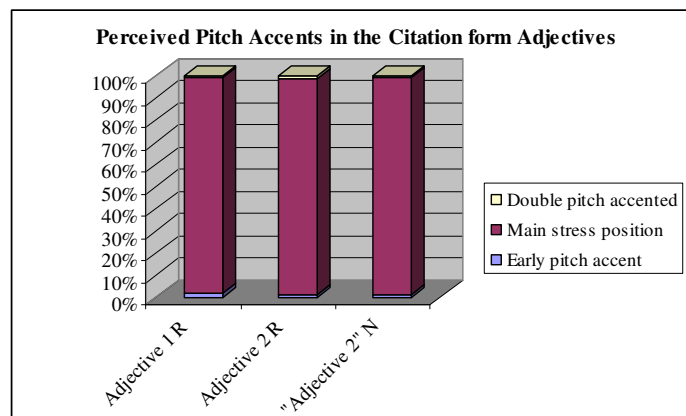
In the next section we try to find acoustic evidence for the perceived early pitch accents.

Figure 50 Percentages of perceived pitch accents on the adjectives in the Recursive phrases (R) (N = 232) and Non-Recursive phrases (N) (N = 238)

a.



b.



#### 4.4.5.2. *Acoustic results*

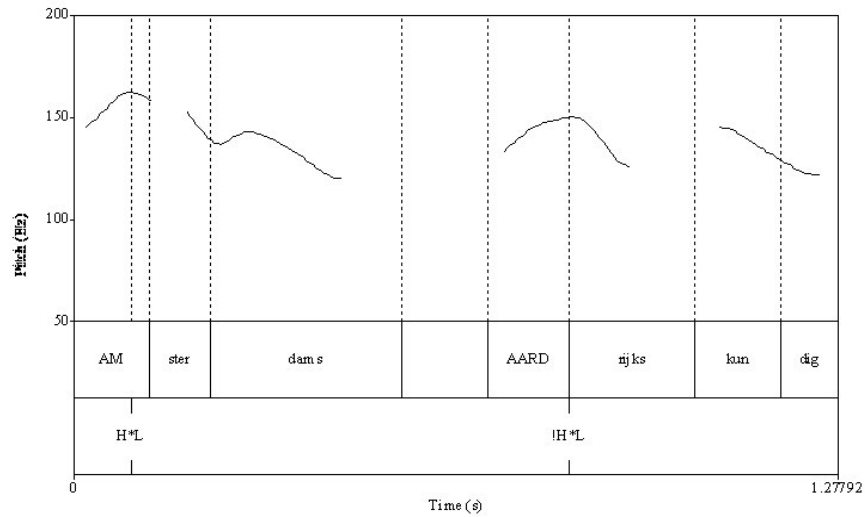
In order to try to underpin our findings with empirical evidence, we measured the values of the correlates maximum and mean fundamental frequency, duration, intensity, spectral balance, and also rhythmic timing.<sup>25</sup> For accent, measuring fundamental frequency actually should suffice, but we did not want to exclude other possible correlates in advance. Figure 51 shows the F0-contours and syllable durations of *amsterdams aardrijkskundig (genootschap)* in the phrase (Figure 51a) and in citation forms (Figure 51b), as realized by the male subject M05. This phrase was perceived as having early accents on both adjectives, whereas the words in citation forms were accented on the main stress position. This is reflected in these F0-contours. The accented syllables are capitalized in the textgrids.

---

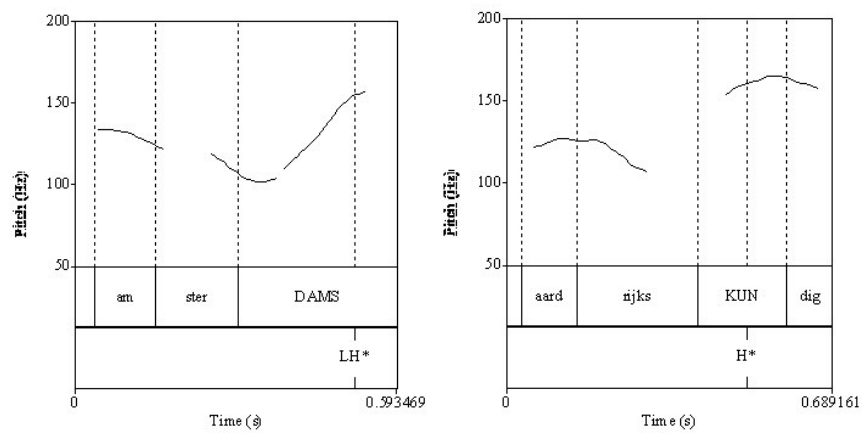
<sup>25</sup> For the recursive NPs the analyses of pitch, duration, and intensity were performed only on the second adjectives, which were the same adjectives as in the non-recursive NPs. Rhythmic timing was analyzed separately from the more conventional correlates. We will report on these rhythmic timing analyses further on in this section.



Figure 51a. F0-contour of the phrase *àmsterdams àardrijkskundig* (*genóotschap*)



b. F0-contours of the words *amsterdams* and *aardrijkskundig* in citation form



In order to look for the correlate of perceived accent in the phrases, we performed a Multivariate Analysis of Variance (MANOVA) on the phrasal adjectives. We investigated whether the

perception of an accent on a certain syllable depended on the height of (one of) the acoustic correlates duration, fundamental frequency, intensity or spectral balance of that same syllable, either as a bundle of correlates, or as a single correlate. This turned out mostly to be true for the main stress syllable, however not for the early syllable. T-tests comparing the phrasal correlates with those of the individual words also showed no significant differences for the early syllables, but some significant differences for the main stress positions.

These first findings suggested that we should shift our focus from the acoustics of the early syllable to the influence of the main stress syllable acoustics on the perception of the early syllable. Therefore we performed another MANOVA, but now with the perceived accent on the early syllable compared with the acoustic values of the main stress syllable. Indeed, where an accent was perceived on the early syllable, the acoustic correlates of the main stress syllable show lower mean values, except for intensity, as can be seen in Table 26 and Figure 52. The correlates that do show lower main values are fundamental frequency (pitch) and duration. Notice that the high standard deviations for the fundamental frequency are partly caused by the fact that we analyzed male and female speakers together.

This may seem a strange finding, meaning that listeners perceive a change on one syllable, while the change in reality had occurred on another syllable. However, Horne (1990), Grabe and Warren (1995), and Vogel, Bunnell and Hoskins (1995) also find evidence for accent deletion on the basis of phonetic data. The main stress position under stress shift represents a deleted pitch accent, while the secondary stress is unchanged. The finding in this chapter of course strongly supports this view. Gussenhoven (1991, 2005) accounts for non-peripheral accent deletion in phonological phrases phonologically. According to his proposal the distribution of pitch accents in English is determined by the interaction of constraints on lexical accent working on two lexical levels, and postlexical rhythmic readjustments (*cf.* section 4.4.5.2).

The fact that the perception of an accent on the early syllable depends on the deletion of an accent on the main stress syllable, is indicative for the relativity of our perception: the removal of an

accent on one syllable has the effect of a perceptive accent on another syllable, while no acoustic accent cues may be present on that syllable. Table 27 shows the result of the MANOVA: all the differences but intensity are significant.

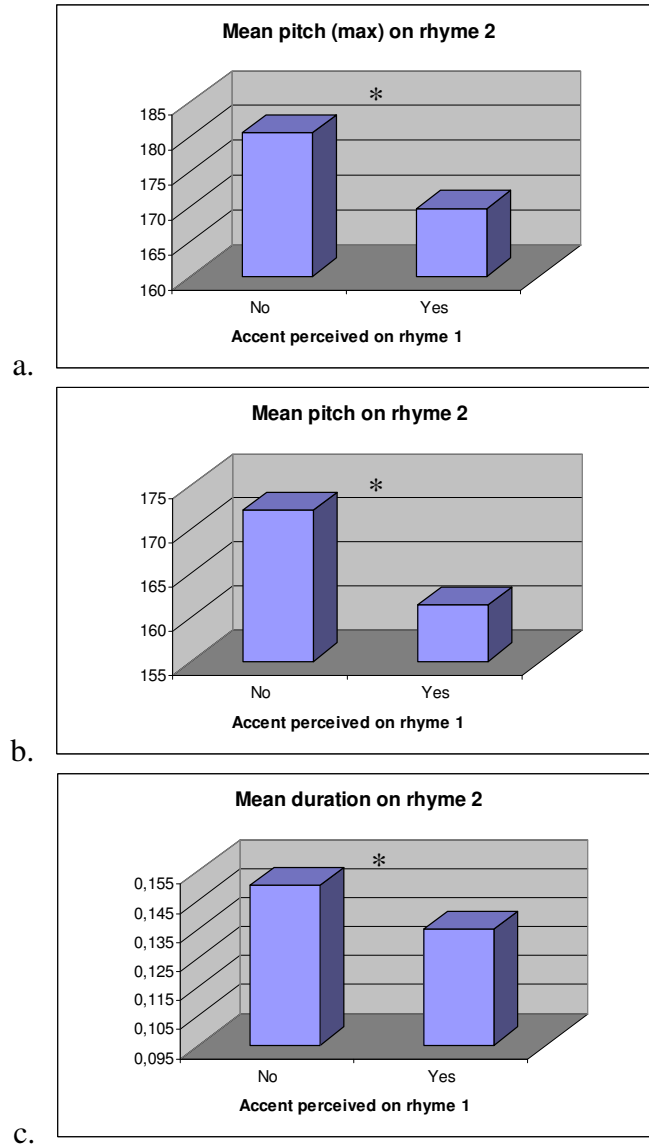
Table 26 MANOVA: Descriptive Statistics

Acoustics on main accent position	Accent perceived on early syllable	Mean	Std. Dev.	N
pitch max	No	180.5941	52.87693	266
	Yes	169.6874	46.29735	184
pitch mean	No	172.1913	49.80910	266
	Yes	161.4167	44.54887	184
duration	No	.1500	.06776	266
	Yes	.1357	.05071	184
intensity	No	72.2209	5.07711	266
	Yes	72.1551	6.08775	184

Table 27 MANOVA: dependencies between perceived early accents and absence of acoustic accents on the main accent position

Source	Dependent Variable (acoustics of rhyme 2)	F	Sig.
Perceived accent on rhyme 1	pitch max	5.115	.024
	pitch mean	5.542	.019
	duration	5.869	.016
	intensity	.016	.901

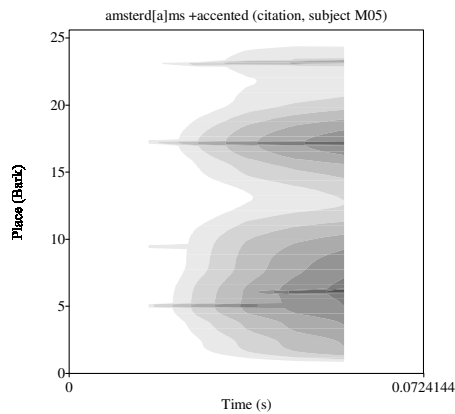
Figure 52 Acoustics on main accent position



We also performed an impressionistic investigation on spectral balance in a small part of the data. (For the description of this correlate, see Chapter 3.) It appeared that spectral balance did not contribute to the effect of accent. In Figure 53a,b,c we show the cochleagrams and the loudness distribution over the spectrum of the main stress syllable of the word ‘amsterdams’ with an accent on that syllable in citation form, and without an accent in phrasal form. The spectral balance in these pictures shows the expected higher energy in the higher frequency regions in the accented case. However, in most other cases the differences were unpredictable. So we assume spectral balance not to contribute to the perception of (early) accent.

Figure 53 Spectral balance on main accent position

a. Cochleagram of +accent [ɑ]



b. Cochleagram of -accent [ɑ]

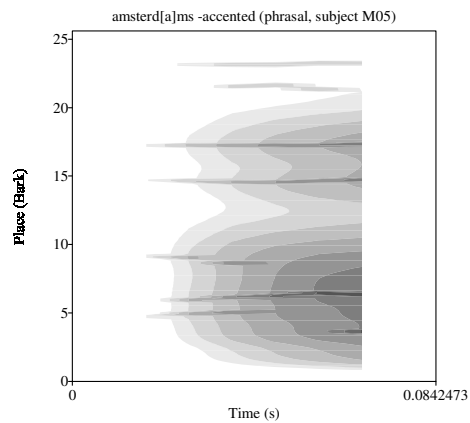
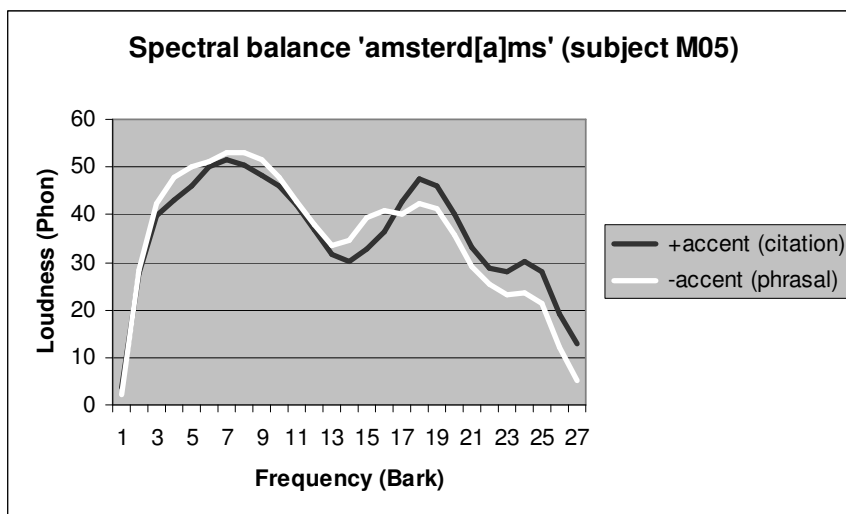


Figure 53c. Loudness within the spectrum



Subsequently, we were interested in the differences between the main stress syllables of the phrases and the words in citation form. We subtracted the values of the correlates of the citation form adjectives from the values of the corresponding phrasal adjectives. An outcome smaller than zero meant that the main stress syllable of the citation form word had higher values than those of the phrasal words, hence the main accent syllable of the phrasal word should have been de-accented. Outcomes smaller than the Just Noticeable Differences (Rietveld and Van Heuven 1997 and references therein) are counted as having equal values in citation form and in the phrase.

With a Pearson Chi-Square test we compared the outcomes of the main stress syllables with the perceived accents on the early syllables. The results from our pilot experiment (138 items) showed that the de-accenting of the main stress syllable did in fact result in the perception of an accent on the early syllable. This is particularly true for the correlates mean pitch ( $\chi^2$  (df2) = 13.056,  $p < 0.001$ ) and duration ( $\chi^2$  (df2) = 6.891,  $p < 0.05$ ).<sup>26</sup> Maximum pitch gives just a

<sup>26</sup> Note that many words that were not perceived as early accented also have low values on the main stress syllable. This seems to contradict the other findings. Possibly, this is due to the fact that the early syllables of many of the words in

marginally significant result and intensity rather does the opposite from the other correlates and the dependency is not significant. Intensity may be not a reliable correlate, due to testing circumstances like the variable distance of the subject's mouth to the microphone. The dependencies between accent and the correlates mean pitch and duration are plotted in Figure 54. The data of the full experiment did not confirm these last findings with significant results. Nonetheless, the mean values of the phrasal data alone do point to the dependency between the acoustics of the main stress syllable and the perception of an accent on the early syllable, as shown above.

As mentioned in the introduction of this section, we also analyzed rhythmic timing. Quené and Port (2003) found that the perception of early pitch accents depends on equal spacing, i.e. rhythmic timing. Note that this is a different kind of rhythm than counting syllables as in the eurhythmy rules. We looked for the same evidence in a subset of our data. First we compared, by means of a t-test, the two inter-accent intervals between the perceived early accents on adjectives 1 and 2 and the main stress accent on the noun. These were found to be similar ( $t(df30) = -.307, p > 0.05$ ). Figure 54a shows the equality of these two intervals, as compared to the intervals between the non-accented syllables in adjectives 1 and the early accented ones in adjectives 2.

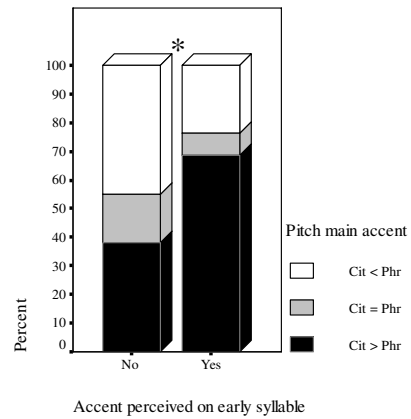
This points to the same conclusion as Quené and Port's (2003), that the perception of early pitch accents is caused by rhythmic timing. However, in the foregoing paragraphs we have shown that, as opposed to our data in the fast speech experiment (chapter 3), the perception of an accent does not entirely depend on rhythmic timing alone, because the acoustic correlates pitch and duration of the main stress position – de-accenting – also influence the perception of early accents.

---

citation form have relatively high acoustic values, just not high enough to be perceived as an accent, relative to the main accent syllable. These high early syllables reflect the fact that words in citation form are phonological phrases by themselves. Normally, phrases consisting of a single word are not signaled by early accent placement, but nevertheless this may have caused the high 'Cit > Phr'-bar in the chart in Figure 54 where no early accent was perceived in the phrases.

Figure 54 Acoustic values Pitch and Duration of the words in citation form subtracted from the values of words in a phrase (pilot experiment, 138 items)

a.



b.

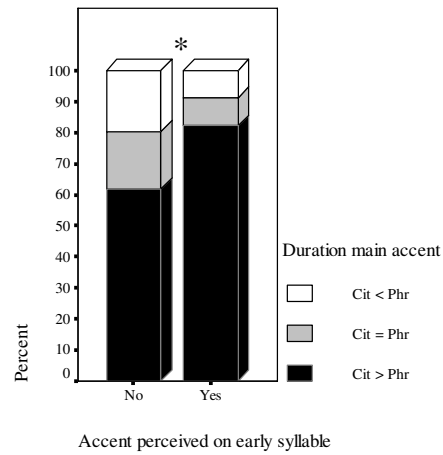
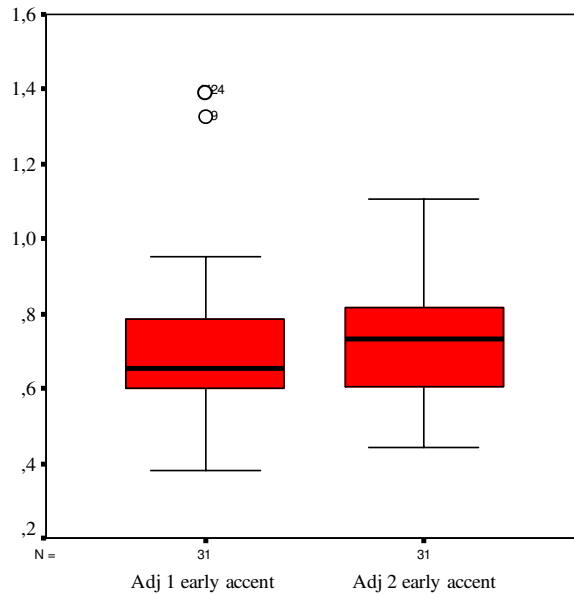


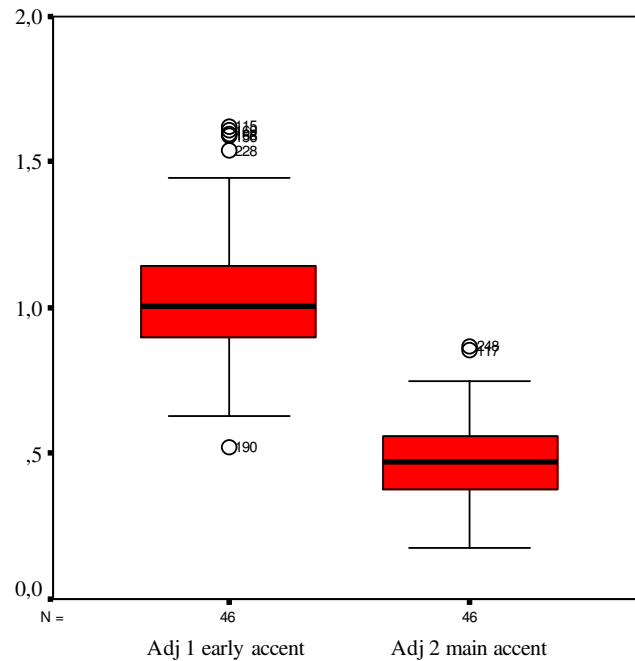


Figure 55a Boxplot of inter-accent intervals between the early accents on adjectives 1 and 2.



Moreover, if timing caused the perception of accents, this should also hold for non-shifted accents. We compared the inter-accent intervals between the early accented syllable of adjective 1 and the accented main stress position of adjective 2 on the one hand, with the interval between this last accent and the accent on the noun on the other. These intervals appeared to be significantly different ( $t(df44) = 13.487$ ,  $p > 0.001$ ), as the boxplot in Figure 55b shows. We must conclude that more is involved than rhythmic timing only. On first sight, the Equal Spacing Constraint seems to have a strong influence on the perception of early accents, but we showed that this is not the case between early and non-shifted accents.

Figure 55b Inter-accent intervals between the early accents on adjective 1 and the accents on main stress position on adjective 2



#### 4.4.6. Phonological analysis

For the phonological analysis of the data, we will give an OT analysis, with a fixed constraint ranking. The optimal output must be the most observed candidate. Suboptimal alternatives, however, can emerge as variants. As in the previous chapter, variation forms are modeled as local optima with respect to a neighborhood structure on the set of candidates, in terms of the optimisation algorithm of Simulated Annealing (Bíró 2005, to appear). In the previous chapter we described Bíró's Annealing Simulation for our fast speech data, and this simulation proved to be successful. We did not perform an Annealing Simulation for the current data, but it is possible that a

simulation would come up with similar percentages (Bíró, personal communication). It should be tested in future work, however.

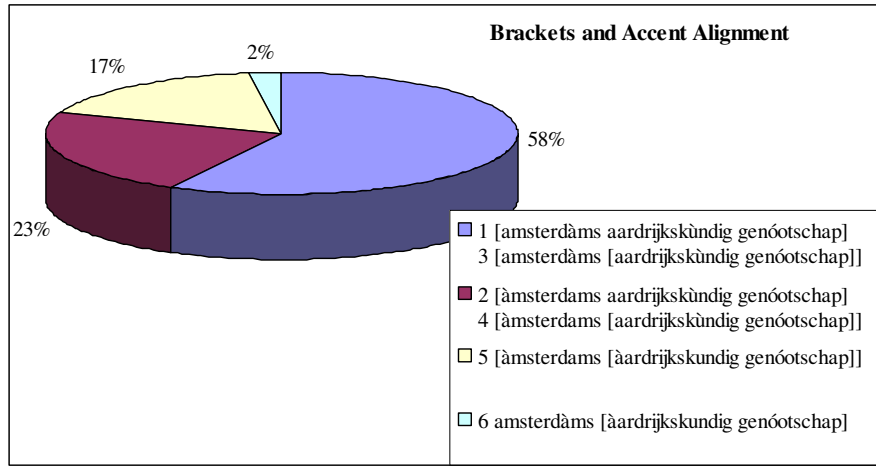
The constraints used in Table 28 are WRAP-XP, OUTPUT-OUTPUT CORRESPONDENCE (O-O CORR), LEFTALIGN, and ALIGN-XP. The constraint WRAP-XP demands that the elements of an input morphosyntactic constituent of type XP must be contained within a Phonological Phrase in the output representation (Truckenbrodt 1999). WRAP-XP also has the effect of deleting the phrase boundaries after each maximal XP (Horne et al. 1999: 73). O-O CORR, as we described in Chapter 3, demands that all variants of the same individual word parts have the same elements and prosodic structure. LEFTALIGN aligns an accent with the left edge of a phonological phrase, and ALIGN-XP is a constraint which aligns syntactic and prosodic boundaries (Selkirk 2000). Horne et al. (1999) assume a similar analysis. They observe:

“The ALIGN-XP constraint makes each maximal phrase category a prosodic phrase, and can come into conflict with WRAP-XP, because the latter constraint has the function of ‘wrapping’ an entire utterance into one prosodic phrase, whereas ALIGN-XP, unconstrained, would lead to an utterance having as many prosodic phrases as there are maximal phrase categories.” (Horne et al. 1999: 73).

This is exactly the way we analyse our data. The most important requirement for getting prosodically recursive phrases is the interaction between a constraint that ‘places the brackets’, i.e. that provides the right prosodic structure, and a constraint that aligns accents to these brackets. This is realized by WRAP-XP, ALIGN-XP and LEFTALIGN in Table 28. In Figure 56 we see that the percentages are as we would predict, following the constraint ranking in Table 28. On the other hand, all possible accent structures occur, including the less optimal ones.

Table 28 OT ranking (the output numbers refer to the numbers in Figure 56)

	WRAP-XP	O-O CORR	LEFTALIGN	ALIGN-XP
1			*	*
2		*!		*
☞ 3			**	
4		*!	*	
5		*!*		
6	*!			*

Figure 56 Observed percentages of different accent structures for the syntactically recursive NPs<sup>27</sup>

Output 3, [amsterdàms [aardrijkskùndig genóotschap]], is the optimal realization of the ranking in Table 28. It has the same realization as output 1 with regard to the accentuation pattern, which

<sup>27</sup> The percentages in Figure 56 are somewhat different from the percentages mentioned in the text. This discrepancy is due to the fact that for this graph we did not count the items with non-flexible accent patterns in Adjectives 1. We were most concerned with the patterns of Adjectives 2. Items with double accents are left out here as well for the sake of clarity.

indeed occurs in a vast majority of the cases. However, output 3 has an (inaudible) recursive phrase structure, while output 1 has a flat phrase structure. In Simulated Annealing, Outputs 2/4 and 5 can probably be variants if they form local optima. Coetzee's model (2004) would also predict the variants as suboptimal alternatives, with the cut-off point, based on the observations, set between WRAP-XP and O-O CORR. In a Stochastic model they could also be variants if LEFTALIGN and ALIGN-XP overlap with O-O CORR, in combination with a high noise level. Output 5 is the only candidate with an audible recursive phrase structure, which is due to the early accents. Output 6 occurs only in 2% of the cases and can therefore be seen as noise. It violates the highest-ranked constraint WRAP-XP.

The outputs 1 and 3, and 2 and 4, are not distinguishable auditorily. Our OT analysis happens to prefer recursive structures, however. Although our data show that a recursive phrase structure with an aligned accent pattern is not the preferred option, the vast majority of the data possibly does have a recursive phrase structure. If our analysis is right, this implies another argument against the Strict Layer Hypothesis.

Simultaneously with our analysis, Gussenhoven (2005) independently proposes a recursive analysis of multiply premodified NPs in English as procliticized phrases. He claims that the distribution of accents is determined by the interaction of lexical and post-lexical accent rules, with two lexical levels and one post-lexical level. The outputs of Level 1, where accents on prenuclear and nuclear feet are assigned and where Level 1 affixation is arranged, are input to the rules of Level 2, such as the COMPOUND RULE and Level 2 affixation. Level 2 outputs in their turn are input to a postlexical level, where phrases are formed. At the postlexical level Alignment constraints regulate the phrasal accent distributions. Differences in accent structures are, in this account, caused by the different accent distributions that leave the lexicon.

Gussenhoven claims that cyclic derivation, with OUTPUT-OUTPUT CORRESPONDENCE, does not work, because such a derivation would wrongly predict accentuation at the right boundary (output from Level 2) between an embedding premodifier and the head, as on -teen in *[[FIFTEEN] [Japanese conSTRUCTIONS]]* (capitals following Gussenhoven). Essential here is the constraint  $\text{ALIGN}(\phi, T^*, \text{RT})$

(AlignRight), which aligns accents to the right edge of phonological phrases. The presence of the right boundary in combination with  $\text{ALIGN}(\emptyset, T^*, \text{RT})$  causes cyclicity to predict the wrong accent pattern. The solution Gussenhoven proposes is that the premodifiers are clitics, with a lefthand  $\emptyset$ -edge without at the same time creating a righthand  $\emptyset$ -edge. This means that *fifteen* does not embed *Japanese constructions*, but extends it. That is why he calls these phrases ‘procliticized phrases’. The absence of the right edge deprives  $\text{ALIGN}(\emptyset, T^*, \text{RT})$  of its erroneous influence. Besides the alignment constraints, the constraint NOREMOTECLASH, a constraint that demands that accents are more than one unaccented syllable apart, is needed (firstly for this data) to avoid *-nese* in *Japanese* being accented. In fact, this constraint bans alternation, which makes it an odd constraint with respect to the universality of OT constraints.

If we apply this analysis to our data, we come up with output 5 (*cf.* Figure 56) as the optimal candidate, while our more frequently observed outputs 1/3 and 2/4 crash on  $\text{ALIGN}(\emptyset, T^*, \text{Left})$ . Candidate 1 in Table 29, which corresponds to our winning candidates 3 and 1 in Figure 56, would be the worst candidate in Gussenhoven’s account. We illustrate this in Table 29a,b.

Table 29a Output candidates of our data using Gussenhoven’s account, input is output from Level 2

Input:	AMsterDAMS	AARDrijksKUNdig	geNOOTschap
Candidates:			
1	[amsterDAMS	[aardrijksKUNdig	geNOOTschap]
2	[AMsterdams	[aardrijksKUNdig	geNOOTschap]
3	[AMsterdams	[AARDrijkskundig	geNOOTschap]
4	[amsterDAMS	[AARDrijkskundig	geNOOTschap]
5	[AMsterDAMS	[AARDrijksKUNdig	geNOOTschap]
6	[AMsterDAMS	[AARDrijkskundig	geNOOTschap]
7	[AMsterdams	[AARDrijksKUNdig	geNOOTschap]
etc.			

Table 29b OT analysis of our data using Gussenhoven's account

	Dep-IO(Accent)	ALIGN( $\phi$ , T*, Rt)	ALIGN( $\phi$ , T*, Left)	NOCLASH	NOREMOTECLASH	MAX-IO (Accent)
1			*!*			**
2			*!			**
☞ 3						**
4			*!	*	*	**
5				*!	***	
6				*!	**	*
7					*!	*
etc.						

Output candidates 1, 2, and 4 violate ALIGN( $\phi$ , T\*, Left) because they have no accent on either *am-* or *aard-*. This is already fatal for these candidates, and candidate 1 violates it even twice. NOCLASH is violated by candidates 4, 5, and 6, because two adjacent syllables are accented. Candidate 4 and the candidates with multiple accents per word (5, 6, 7, etc.), fail on NOREMOTECLASH, because only one unaccented syllable intervenes between the accented ones in one or more places. Notice that candidate 7, and probably more of such multiply accented phrases, falls out later than candidates 1 and 2, while these are observed in our data, and they are the preferred outputs according to our analysis in Table 28.

What is more, standard OT only allows one derived level, namely the output. Revising OT to a version in which different levels each have their own constraint ranking (*cf.* Kiparsky 1982a,b, D.B. den Ouden 2004) would reduce the restrictiveness of the theory with a proliferation of the same kind of constraints that work at different levels. Furthermore, a more modular model would prohibit the interaction of different influences on outputs and consequently reduce the possibility of accounting for so-called ‘conspiracies’ of different

influences in phonology, which turned out to be one of the great merits of OT.

Gussenhoven proposed this analysis for English phrases. For Dutch, it appears not to predict the right results, and the same holds for French (Gussenhoven personal communication). Our results also contradict the findings of Visch (1989), who concludes that in Dutch the tendency to choose early accent placement over O-O CORR is much stronger than in English. Because our own analysis can cope with our Dutch data better, we will stick to our straightforward and simple analysis.

#### **4.5. Conclusion**

The results of this experiment show that the prosodic recursion hypothesis holds: recursion does exist in prosody. We have shown that phonological phrase boundaries are often signaled by early accent placement, though not always. Our data also show that prosodic structure is less linear than assumed in the Strict Layer Hypothesis and derived hypotheses. In 22% of our data, the second adjective in a syntactically recursive noun phrase was perceived as having an early accent, and therefore it is reasonable to assume that these phrases were prosodically recursive as well. What is more, the OT analysis suggests that possibly a vast majority of our data may have a recursive phrase structure, though this structure is inaudible without the early accents. The analysis has a clear preference for this structure. In other words, the results of our experiment must be understood as additional evidence for a more prominent place for recursion in phonology. It would have been peculiar, after all, if recursion were not to exist in prosody, while it is currently seen as the most characteristic feature of linguistic syntax (*cf.* Hauser et al. 2002), and it is found in all kinds of structures in the world, such as nature itself, visual art, and music, as we showed in section 4.2. One could say that syntax consists of tacit cognitive structure, while phonology deals with the cognitive structure of physical behavior, which sets it apart. But especially the finding that music, as another



physically performed cognitive behavior, shows recursive structure, is an argument for assuming recursion in phonology as well.

Yet the results also show that the embedded phonological phrases, e.g. *aardrijkskundig genootschap*, do not behave identically to the maximal phonological phrases, e.g. *Amsterdams aardrijkskundig genootschap*, in the sense that the maximal phonological phrases are early accented significantly more often. Recall that the OT analysis showed that these seemingly non-recursive structures preferably do have recursive phrase structure. The difference is that the early accent placement, as a signal for phrase structure, is a less strong tendency than assumed. Clearly, there is a lot of optionality involved, not only for the subjects to apply the process of early accent placement, but also with regard to the recursivity of prosodic structure itself. The results confirm the observation that there is no one-to-one mapping from syntax to prosody, because optionality in syntactic structure would not be possible.

Another conclusion which came out of our analyses is that the perception of an early accent is often based on the de-accenting of the main stress syllable (see also Horne 1990, Gussenhoven 1991). Apparently pitch accent perception is relative: the absence of an accent on the expected position is interpreted by the listener as an accent on the next accentable position to the left. This is another indication that listeners base their perception not only on the acoustic signal alone (*cf.* Chapter 3); some strategy for retrieving the linguistic structure of the utterance must also play a role.

The propensity to get an early pitch accent does not depend on the number of intermediate syllables, as was assumed in the Eurhythmy rules (Hayes 1984). It does depend on rhythmic timing, however. Again, we find the ideal interval of ~ 600 ms between accents. This finding confronts us with the question whether early accent placement should still be assumed as a phrase-marking device or, conversely, as a purely rhythmic phenomenon. Although the finding that rhythmic timing plays an important role in early accent placement seems to weaken the assumption that early accent placement is a phrasing phenomenon, we cannot reject this assumption. The fact that it only occurs in phrases, and not in the same adjectives as individual words, is still an important indication

that it is a phonological structuring device. Besides that, we showed that the Equal Spacing Constraint did not apply between early and non-shifted accents.

In sum, we can conclude that we found strong indications for recursion in phonology, on the basis of auditory as well as on the basis of acoustic pitch and duration data.

## Chapter 5

# Speaking in Minor and Major Keys

### 5.1. Introduction<sup>28</sup>

The prosodic phenomena discussed in the foregoing chapters were all instances of linguistic prosody. Prosody, however, also involves extra-linguistic characteristics, such as emotion. In this chapter we investigate whether or not differences in emotional speech are characterized by different modalities.

In music the difference between sad and cheerful melodies is often indicated as a difference between a minor and a major key. Our main objective is the identification of analogous interval differences in the pitch contours of emotional speech in Dutch. It is obvious that the range in the pitch contour of sad speech is much smaller than the range in cheerful speech, but do we also speak in a minor key when we are sad and in a major key when we are happy?

As we described in Chapter 1, Lerdahl and Jackendoff (1983) and Gilbers and Schreuder (2002) among others observe that intonation patterns in speech and melodies in music have a lot in common. One of the linguistic functions of intonation patterns and melodies is to mark boundaries. Differences in pitch movement can cause different meanings.

In order to investigate emotional intonation, we recorded and analyzed the performances of five professional readers reading passages from A.A. Milne's *Winnie the Pooh* in Dutch (1994, 1995). We are interested in the sad character Eeyore and the happy, energetic Tigger. Although we do not find modality in the pitch contours of all speakers, we do find intervals between tones

---

<sup>28</sup> A version of this chapter also appeared as Schreuder, Van Eerten, and Gilbers (2006). The data were gathered by Laura van Eerten (2004), and she is also responsible for part of the analyses.

indicating minor modality exclusively in Eeyore passages and intervals indicating major modality exclusively in Tigger passages.

This chapter is organized as follows. In section 5.2 we outline the theoretical background; in section 5.3 we describe the method of our experiment and in section 5.4 we give the analysis and the results, which are discussed in section 5.5.

## 5.2. Theoretical Background

The difference between sad and cheerful music is often described as a difference between a minor and a major key, although in some instances composers play around with the notions of major and minor modality, which may result in cheerful music in a minor key, or sad music in the major key. The scale in western tonal music is divided into twelve steps, also called ‘semitones’. Typical of the minor modality is that it features chords that are characterized by a distance of three semitones between the tonic and the (minor) third, whereas chords in the major modality feature a distance of four semitones between the tonic and the (major) third. This difference in thirds is the main factor for the perception of mood in music.

Figure 57 Keyboard

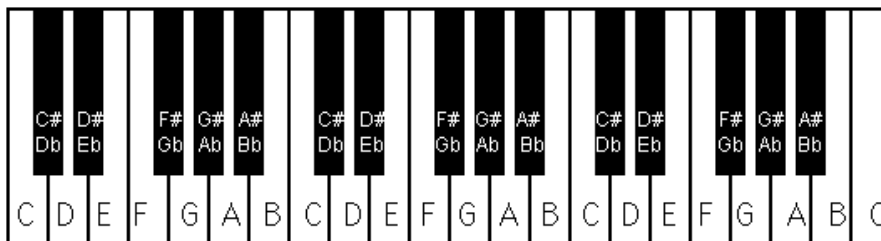


Figure 57 shows the keys of a keyboard instrument. The distance between C and C#, for instance, involves one semitone; the distance between C and D two semitones. Thus, a minor third is constituted by C and Eb and a major third by C and E.

Each note has its own frequency. For example, the concert A is 440 Hz. A' one octave higher has a double frequency: 880 Hz; A one

octave lower has a frequency of 220 Hz. Within the octave, A and A' are twelve semitones apart: five black keys and seven white keys in Figure 57. The frequency ratio between two semitones is equal. It is the twelfth root of two, which is approximately 1.0595. Table 30 shows frequency values of each note.

Table 30 Note frequencies in Hz

C	65.4 Hz	C	130.8 Hz	C	261.6 Hz
C#	69.3 Hz	C#	138.6 Hz	C#	277.2 Hz
D	73.4 Hz	D	146.8 Hz	D	293.6 Hz
D#	77.8 Hz	D#	155.6 Hz	D#	311.2 Hz
E	82.4 Hz	E	164.8 Hz	E	329.6 Hz
F	87.3 Hz	F	174.6 Hz	F	349.2 Hz
F#	92.5 Hz	F#	185.0 Hz	F#	370.0 Hz
G	98.0 Hz	G	196.0 Hz	G	392.0 Hz
G#	103.9 Hz	G#	207.7 Hz	G#	415.3 Hz
A	110.0 Hz	A	220.0 Hz	<b>A</b>	<b>440.0 Hz</b>
A#	116.6 Hz	A#	233.2 Hz	A#	466.2 Hz
B	123.5 Hz	B	247.0 Hz	B	493.9 Hz

Braun (2001) studied Dutch speech and found out that the majority of the speakers speak according to an internal tuned scale. Cook (2002) and Cook, Fujisawa and Takami (2004) investigate the modality of Japanese emotional speech. Normally, the pitch range of about seven semitones is used in sentences, a fifth. Cook et al. conclude that utterances perceived as having positive affect significantly show major-like pitch structure, whereas sentences with negative affect have a tendency to minor-like pitch structure. The conclusions are based on cluster analyses of the pitch contours of recorded utterances. In these cluster analyses the actual pitch values at every millisecond are rounded down or up to the value of the nearest semitone (*cf.* Table 30).

In this chapter, we present a follow-up to these studies, in which we try to find out whether there are different modalities in Dutch emotional speech. Apart from cluster analyses we will also

investigate sequences of individual notes in scores of emotional speech.

### **5.3. Method**

In order to obtain different emotions in speech, we asked five primary school teachers to read out selected passages in Dutch from A.A. Milne's *Winnie the Pooh*, in which energetic, happy Tigger, and distrustful, sad Eeyore, are presented as talking characters.

The primary school teachers are experienced readers. The two men and three women aged 27 to 32 all claimed to have musical affinity; four of them played an instrument. They all read out the same passages, which were recorded on hard disk as wav-files and analyzed using the software programs CoolEdit 2000 and PRAAT (Boersma and Weenink 1992-2006).

The passages in which Tigger and Eeyore speak were extracted and concatenated into twenty files each varying from 8 to 53 seconds. The pitch information of these files was measured every ten milliseconds using Praat. In this way we obtained sequences of frequency values representing the pitch contours. Comparison to the original pitch contours revealed a great similarity. Therefore, we decided that this sample rate of ten milliseconds was sufficient for our experiment.

Subsequently, we did a cluster analysis of the pitch data in order to find out which frequencies occurred most in each contour. For this cluster analysis we relied on a cluster algorithm in Excel presented in Cook (2002) and Cook et al. (2004). The product of the frequency data was calculated, and assigned to the nearest semitone in an equally-tempered scale, resulting in a semitone power spectrum. In other words, the obtained pitch values were clustered i.e. rounded down or up to the value of the nearest semitone. This normalization procedure resulted in a semitone histogram in which one can read which semitones occur most in the utterance. In this way, we made an abstraction of the real pitch values that can be compared to the abstractions phonologists make when they describe various allophones as the realizations of one and the same phoneme. As Cook (2002) remarks, it might be more valid to normalize to the speaker's dominant pitches above the tonic, instead of to the musical

equally-tempered scale, and then study the interval substructure. This would probably lead to somewhat different results, but it would also complicate the analyses.

Furthermore, we converted the pitch contours of the stories into musical scores, to account for intervals in sequences. The aspect of time may be an important property in the analyses of modality.

## 5.4. Analyses and results

### 5.4.1. Cluster analysis

Cook et al. (2004) identify the musical modality of Japanese speech on three peaks in the cluster analysis, because musical modality is based on triads. Nooteboom and Cohen (1995: 157, 162-163), however, claim that the range of Dutch intonation moves between two perceptively relevant declination levels in contrast to the three levels of English intonation. Indeed, most of our graphs show one or two peaks. There are only two graphs with three peaks. Therefore, we decided to indicate the modality on the occurrence of intervals of thirds in the graphs. If the interval between peaks concerns a minor third, we indicate the modality of speech as minor; if the interval concerns a major third, the modality is considered to be major.

Inspection of the cluster analyses shows that not all graphs contain more than one peak. In other words, in graphs with just one peak the modality cannot be determined. These one-peak graphs were found in eight of our twenty sound files. In contrast to tonal music, which usually has a major or minor modality, speech can be neutral.<sup>29</sup> In five cases the peaks are too far apart to decide on the modality. If the peaks constitute a fifth, for example, one cannot determine the modality. This does not immediately imply that all these instances are counterexamples, they are just indecisive. Seven cases remained for analysis.

---

<sup>29</sup> Music with a neutral modality does occur, however. Metal music, for instance, frequently uses so-called power chords, which consist solely of the tonic and the dominant. Without triads, no modality can be derived. Moreover, one can think of music without chords, with a melodic line with intervals of e.g. only fourths and fifths. This is a rare phenomenon in music, while it seems to be a normal option in speech.

Our analyses confirm our hypothesis. The major modality is exclusively found in sound files of Tigger stories in which thirds were observed, whereas the minor modality only appears in sound files of Eeyore stories. We conclude that Tigger speaks in a major key and Eeyore in a minor key in these cases.

Figure 58a shows a cluster analysis example of the raw data of Tigger as performed by subject HJ. The x-axis presents the pitch values in Hertz and the y-axis depicts the number of occurrences of a certain pitch value in the sound file. The frequency range is large, from 87 to 406 Hz.

Figure 58a Tigger in major; cluster analysis

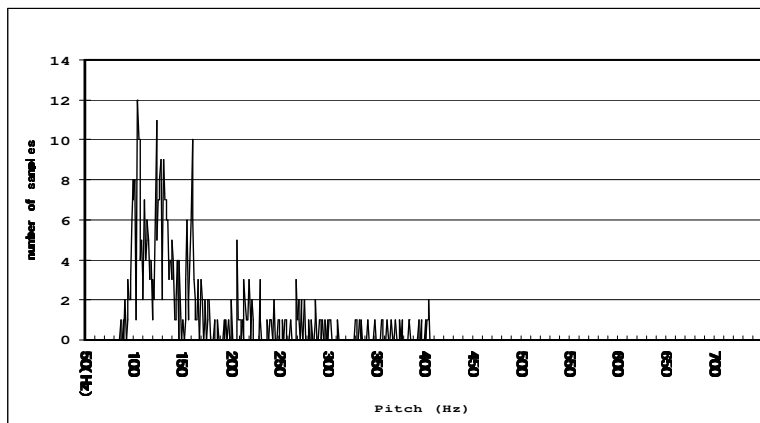




Figure 58b Tiger in major; semitones

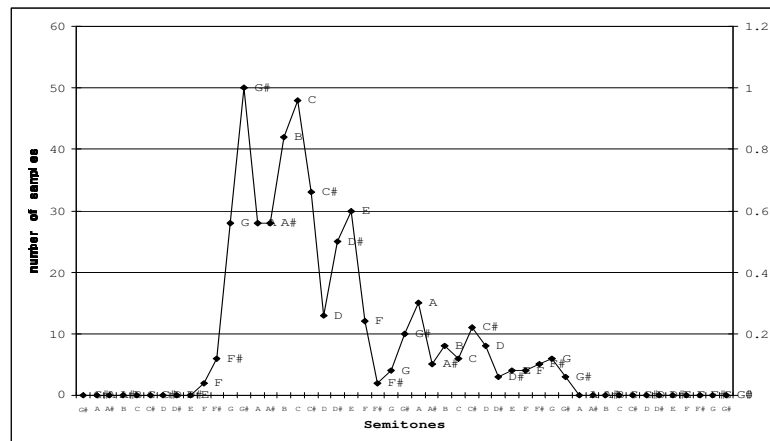


Figure 58b shows the same fragment as Figure 58a, this time clustered in semitones. The figures were obtained using Cook's cluster algorithm macro in Excel. On the x-axis abstractions (musical phonemes) of the real frequencies (musical allophones) are depicted as musical notes. On the y-axis we show the number of samples for each note. Our analyses are based on the semitone graphs, such as the one in Figure 58b.

Figure 58b is one of the few graphs that show three peaks. From left to right the first two peaks are on the notes G# and C. The distance of four semitones between these notes constitutes a major third. The following peak in the graph is at the note E which also constitutes a major third with the preceding C. G# and C form an inverted major third together. Tigger, as spoken by the male subject HJ, is a cheerful character and his speech indeed exhibits the major thirds of a major modality.

Figure 59a shows the clustered data of the same subject HJ's interpretation of Eeyore. The frequency range is smaller this time, from 75 to 200 Hz. In comparison, the frequency range of Tigger was from 87 to 406 Hz. The peaks are also located in lower regions in comparison with Tigger.

Figure 59a Eeyore in minor; cluster analysis

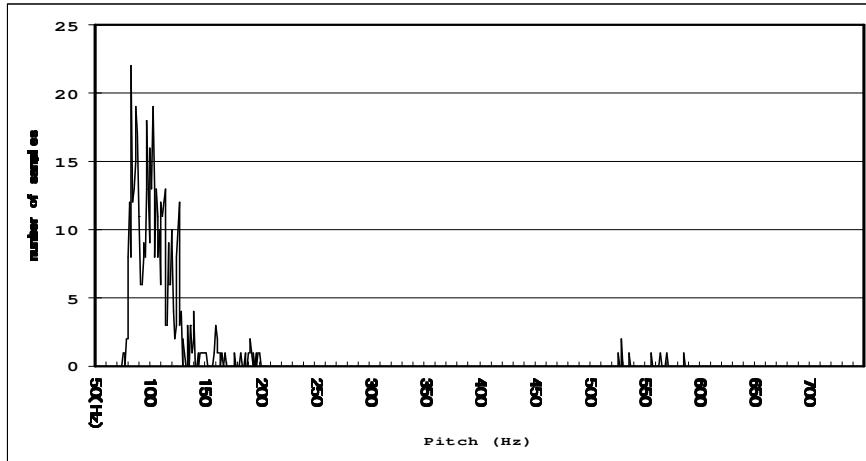
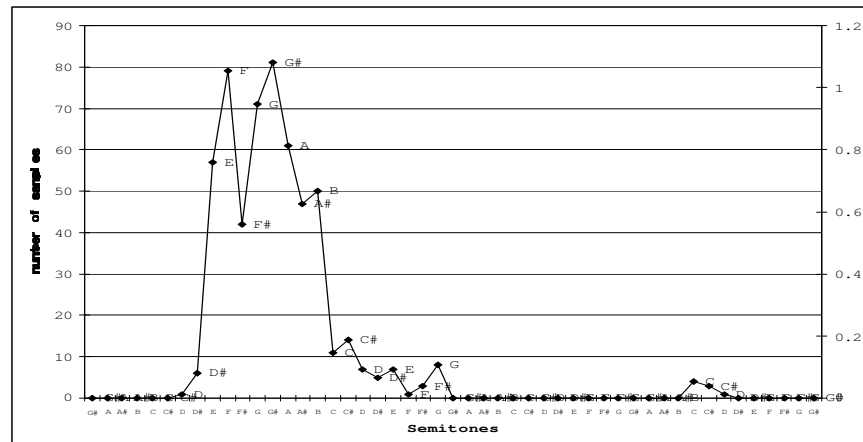


Figure 59b shows the same fragment clustered in semitones with two peaks on, respectively, F and G# (or *Ab*). The distance between the peaks is three semitones, in other words a minor third: Eeyore speaks in a minor modality.

Figure 59b Eeyore in minor; semitones



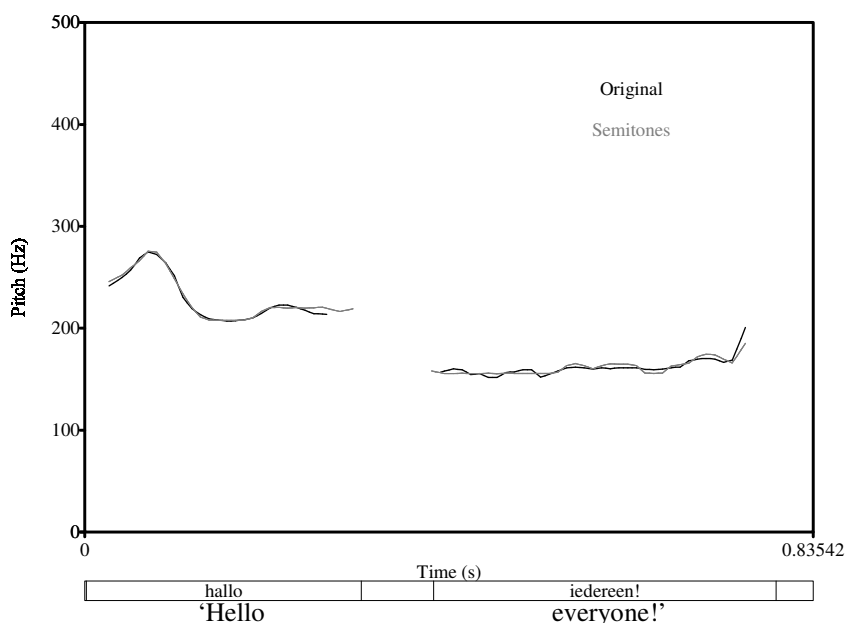
#### 5.4.2. Musical scores

The cluster analysis ignores absolute intervals in time. In other words, the result is not a kind of musical score of speech. Actually, we do not know whether peaks on, for instance, C and E constitute a major third or an inverted augmented fifth. Cook (2002) justifies his choice by claiming that it is unlikely that simply an alteration in the sequence of pitches that conveys positive or negative affect could transform a minor mood into major, or vice versa (Cook 2002, p.118). In music, however, the same melody can cause different moods depending on the chord structure of the song. For example, if a phrase in the key of C is repeated, whilst the chord progression changes to A minor, which is the parallel of C, the mood may change from cheerful to sad.

Therefore, we incorporated time as a factor, which may lead to more reliable results. We did this by using the following formula in Praat:  $2^{\wedge}(\text{round}(\log_2(\text{self}/440) * 12) / 12) * 440$ , which works similarly to a vocoder/harmonizer, rounding off automatically all frequency values at semitone value. The formula calculates the twelfth root of two for rounding off all tones to their nearest semitone, using 440 Hz, the concert A, as a reference tone. Figure 60 shows that, although this manipulation does change the original

values, the differences are very small and do not reach a perceptible level.

Figure 60 Pitch contour of the original speech sound compared to the contour rounded off to the nearest semitone values (Tigger)



Consequently, the manipulated pitch objects were resampled to sine waves.<sup>30</sup> We converted these sine waves to MIDI files, using the freeware program AmazingMIDI (1998-2003). MIDI-files can be represented as musical scores by means of e.g. Steinberg Cubase software or Sibelius. In this way, the resulting musical score of a sound file enables us to determine the modality of the speech.<sup>31</sup>

The resulting scores of two stories, the same stories as depicted in the cluster analyses in Figure 58 and Figure 59, are shown in Figure

<sup>30</sup> Paul Boersma made this possible by adjusting the Praat program twice to our demands. For this we are very grateful to him.

<sup>31</sup> Some MIDI-files and combined speech and musical sound files can be listened to on <http://home.planet.nl/~schre537/sounds.htm> or [www.maartjeschreuder.nl](http://www.maartjeschreuder.nl).

61 and Figure 62. These scores are simplified versions, because a pitch contour consists of several ‘glissandos’, while the MIDI-file must sample the tones into distinct notes. We chose to convert the tones into eighth notes, with the result that all notes of one glissando were unified into single chords. From these chords we chose the most prominent note for each syllable sounding in the original pitch contour. For readability reasons, the Tigger score is in the treble clef, while Eeyore spoke in a lower tone region and is therefore set in the bass clef.

Figure 61 Musical score of the same Tigger story as in Figure 58

**Teigetje moet niezen**  
Verteld door HJ

Lucky me A. Triad Ajar

Hal lo... ie de reen!... Ik heb hem he le maal geen op stop per ver kocht. Nie

tes ik moest nie zen en toen stond ik net ach ter le jaar en toen deed ik,

't was geen op stop per, ik moest nie zen!

In this score of the short Tigger monologue we see the same notes stand out as in Figure 58: G#, C and E, but also A and B. A and B do not form thirds with the other notes. The objective of this score was to look whether (prominent) adjacent notes, ideally notes on neighbouring stressed syllables, form thirds in sequence. This, however, is hard to extract from the score in Figure 61, because most intervals between notes in sequences are larger intervals than thirds. Moreover, most phrases appear to be spoken on a single tone. Comparing intervals between different phrases would be wrong, because in the original speech file parts of text intervened between these phrases.

We find some thirds on stressed syllables, however, which appear to be major thirds: the interval G# – E between *lo* and *ie* in *Hallo iedereen* ‘hello everyone’, and the interval C# – A between *ter* and *Ie* in *achter Iejaar* ‘behind Eeyore’. The major part of this score is built upon notes which form major thirds with each other. This gives the ultimate feeling of a major key: a happy, cheerful, and energetic story.

Figure 62 gives the score of the Eeyore monologue. Again we see many Fs and Abs, as in the cluster analyses in Figure 59. The story is longer, and here we are able to identify sequences of thirds between stressed syllables. Examples are Gb – A in the syllables *maak* and *het* in *hoe maak je het?* ‘how do you do?’, and F – Ab in the syllables *één* and *an* in *de één of ander* ‘someone or other’.

Figure 62 Musical score of the same Eeyore story as in Figure 59

**Iejaar voelt zich niet helemaal hoe**  
Verteld door HJ Dramatical Ray Jeuk

Hoe maak je het? Niet zo erg hoe. Ik voel  
me al een hele tijd niet hoe. Wat is er dan mee? Weet je 't ze ker  
? Wat is er dan wel? Dat moet ik zien. Ik ge loof dat je ge lijk hebt.  
Dit ver klaart veel. Al les wordt me nu dui de lijk. De één of  
an der moet 'm ge sto len heb ben. Zo zijn ze. Dank je wel, Poeh,  
je bent een ech te vriend. En dat kun je niet van ie der een zeg gen.

We did not make (simplified) scores of all the stories. The cluster analyses seem to give a good account of the internal relations in the melodies. While the energetic Tigger speaks in a major key, the melancholic character Eeyore expresses himself in a minor key.

### **5.5. Conclusion**

In this pilot study we analyzed clustered frequency peaks in stories in which the happy Tigger and the sad Eeyore were speaking characters, and we derived musical scores of the pitch contours. The results show that in the cases in which we do find intervals of thirds between the frequency peaks, the major modality is always observed in sound files of Tigger stories, whereas the minor modality is observed in sound files of Eeyore stories. Although thirds were only found in a minority of our material, there were no counterexamples in the fragments containing thirds. The derived musical scores of the intonation contours show that at least the minor thirds of Eeyore can also be found in sequences of stressed syllables.

Although speech can be neutral, we found a tendency that a sad mood can be expressed by using intervals of three semitones, i.e. minor thirds. Cheerful speech mostly has bigger intervals than thirds, but when thirds are used, these thirds tend to be major thirds. Strong conclusions cannot be drawn from only one such a small-scale experiment using a new analytic technique. But the evidence presented above is certainly suggestive. At the very least, these results are an indication that the mood of emotional prosody in speech is rather similar to musical modality. Therefore, this could be a promising method for studying emotion in speech. The tendency we found suggests that further investigation of the similarities between music and speech could be fruitful.





## **Chapter 6**

# **Summary, Conclusions, and Future directions**

### **6.1. Summary and Conclusions**

In this thesis I tackled some phonological issues from a musical perspective. In the first chapter we showed that phonological structure has a lot in common with musical structure and that both kinds of structures can be easily described using the method of Optimality Theory. Some of the similarities between speech and music have a psychological or possibly a neurological basis. In both disciplines preference rules for ideal outputs indicate the prominent constituents of every part of the hierarchical structure, and structurally important elements are assigned salience by aligning tonally or durationally strong elements with them. These resemblances led us to the assumption that insights of music theory can help out in phonological issues. Three of such issues were subjects of the experiments in this dissertation.

The first issue was the question whether the influence of a higher speaking rate leads to adjustment of the phonological structure or just to phonetic compression, or maybe just to a different perception by the listener. We found that listeners perceive the words as rhythmically restructured, with shifted secondary stress. Acoustically, however, we could not find any characteristic of secondary stress. What we did find was that the syllables which were perceived as carrying secondary stress were always located around 300 ms before the main stress syllable. This led us to the conclusion that listeners are equipped with an ‘internal metronome’.

The second issue we discussed was the question whether recursive structures exist in phonology, just as they exist in syntax. Notably, recursion is also a well-known structural characteristic in the extralinguistic world: in nature, visual art, and music. But it would be

odd to assume that phonology does not have recursive structures, as is assumed in the Strict Layer Hypothesis (Selkirk 1984). On the basis of the repeated application of early pitch accent placement in syntactically recursive noun phrases, we showed that recursive embedding of phrases does indeed exist in phonology.

The third issue we experimentally explored concerned the question whether differences in emotional speech are characterized by different modalities. In music the difference between sad and cheerful melodies is often indicated as a difference between a minor and a major key. In a small-scale experiment with sad and cheerful stories we found indications that people also speak in a minor key when they are sad and in a major key when they are happy.

In our view, the observation that language and music show so many similarities strengthens the hypothesis that the same structures and principles hold for all temporally ordered behavior. It is the way in which our brain works: our cognitive system structures the world surrounding us in a particular way in order to understand everything in the best way.

The most striking outcome of our experiments on rhythm and recursive phrasing is that sometimes people hear things that are not present in the acoustic signal. In the case of recursive phrasing in Chapter 4, accents were perceived on the early syllable, while the only difference with the words in isolation was situated in the main stress syllable. This means that a change in one syllable may induce the perception of a change in a different syllable. This is also known as de-accenting (Horne 1990, Gussenhoven 1991). The speaker has the facilities to mark the structural entities in his speech. The listener, in his turn, can add such structural markers to his perception of the speaker's message if these markers are absent. Apparently, listeners base their perception not only on the acoustic signal alone; some strategy for retrieving the linguistic structure of the utterance also must play a role.

A similar effect of deceptive perception was found in Chapter 3 on rhythm in fast speech. Secondary stress was perceived in nearly all words, and the places of these secondary stresses were quite consistently perceived. Nevertheless, none of the acoustic measurements could indicate the secondary stresses. We found out that the listener focuses, unconsciously, on certain points in time, using an 'internal metronome', expecting information of some

importance in those positions. He therefore thinks that he hears an accent, even if it is not actually present. This again points to a communicative strategy the listener uses to anticipate the most prominent parts of the message, which he wants to extract from the sound signal.

The reason the listener uses such an anticipatory strategy is probably that human working memory is finite. One cannot focus on the whole signal altogether. Therefore, one has to extract parts of the signal, and one needs a strategy to determine which parts are essential to understanding the message. Since stress and accent are major cues of salience, the fact that listeners report that they perceive stress on a syllable they expect to be significant is therefore a logical consequence.

One can imagine that an equal tuning of the speech rate of speaker and listener would probably lead to optimal communication in both directions. The speaker will then put the most important parts of the message on the temporal locations the listener is anticipating. Regular rhythm, as Quené and Port (2005) show, leads to faster speech recognition; it enhances the perceptability of speech. Here again we find a similarity between speech and music: a regular rhythm in music is also easier to remember, and prominent chord or notes in a melody, like structure-marking accents in speech, are aligned with prominent rhythmic positions. What the studies in this dissertation show is that the idea that salience is signalled by equal timing in speech performance works the other way around too: the listener expects something of importance in the signal and therefore he actually thinks that he hears an accent.

This listener-based perspective is quite new for phonology. Of course, perception studies have always been a very important element in the discovery of phonological structure. However, the finding that an important part of this perception is based on auditory illusions, where the acoustic measurements of the syllables do not coincide with the perceived differences in prominence, is a new insight. In music theory, this perspective has been taken by several musicologists, as we indicated in Chapters 2 and 3. The conclusion that it also holds for speech rhythm shows once more that language and music share the same kinds of processes, which must partly be based on the perception by the listener. The assumption that music

and speech share some cognitive characteristics is therefore a logical one.

## **6.2. Future directions**

In this dissertation we found that part of our communication is based on the subjective experience of the listener, not on an objective representation of the stimulus alone. This has quite some implications for the methods of research in phonology. If the interpretation of the world is influenced by the mind of the perceiver, then empirical evidence must be sought for not only acoustically, but also outside of phonetics, in psychology for instance. In other words, it is not always the case that, '*meten is weten*' as we say in Dutch, which means that 'to measure is to know' does not always apply.

This listener-based perspective, and the communicative strategies that seem to be involved in rhythm and accent perception, are a very interesting subject for future investigation. I hope I will get the opportunity to set up some experiments to look further into these strategies. The main goal would then be to identify the extra-acoustic influences on perception.

## References

- Abercrombie, D. (1965). *Studies in Linguistics and Phonetics*. London: Oxford University Press.
- Abercrombie, D. (1967). *Elements of General Phonetics*. Chicago: Aldine.
- Abraham, G. (1974). *The Tradition of Western Music*. Berkeley, California: University of California Press.
- Allen, G.D. (1975). Speech Rhythm: its Relation to Performance Universals and Articulatory Timing. *Journal of Phonetics* 3: 75-86.
- AmazingMIDI (1998-2003). Araki Software, Japan. [Http://www.pluto.dti.ne.jp/~araki/amazingmidi/](http://www.pluto.dti.ne.jp/~araki/amazingmidi/).
- Anttila, A. (1997). Deriving Variation from Grammar: A Study of Finnish Genitives. In: Hinskens, F. et al. (eds.). *Variation, Change, and Phonological Theory* (pp. 35-68). Amsterdam: Benjamins.
- Anttila, A. and Y.Y. Cho (1998). Variation and Change in Optimality Theory. *Lingua* 104: 31-56.
- Attridge, D. (1982). *The Rhythms of English Poetry*. English series no. 14. Burnt Hill, Essex: Longman.
- Auer, P., Couper-Kuhlen, E., and F. Müller (1999). *Language in Time. The Rhythm and Tempo of Spoken Interaction*. New York/Oxford: Oxford University Press.
- Beckman, M.E. (1986). *Stress and Non-Stress Accent*. Dordrecht: Foris.
- Bezooijen, R. van (2001). Regionale omroepen, ádequate voorzieningen. Vershuift de klemtoon echt steeds vaker naar voren? *Onze Taal* 70 (5): 104-106.
- Bíró, T. (2005). When the Hothead Speaks: Simulated Annealing for Optimality Theory. *Proceedings of the 15th CLIN conference*, Leiden.
- Bíró, T. (to appear). *Simulated Annealing Optimality Theory*. PhD Dissertation, University of Groningen.
- Bíró, T., Gilbers, D., and M. Schreuder (to appear). *Variation (with)in OT*. Ms., University of Groningen.
- Boersma, P. and B. Hayes (1999). *Empirical Tests of the Gradual Learning Algorithm*. Ms., University of Amsterdam and UCLA. Also available as Rutgers Optimality Archive [ROA-348].
- Boersma, P. and B. Hayes (2001). Empirical Tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32: 45-86.

- Boersma, P. and D. Weenink (1992-2006). *PRAAT, Phonetics by Computer*, University of Amsterdam. [www.praat.org](http://www.praat.org).
- Bolinger, D.L. (1965). Pitch Accent and Sentence Rhythm. In: Abe, I. and T. Kanekiyo (eds.). *Forms of English: Accent, Morpheme, Order* (pp. 139-180). Cambridge, Massachusetts: Harvard University Press.
- Bolinger, D.L. (1981). *Two Kinds of Vowels, Two Kinds of Rhythm*. Bloomington: Indiana University Linguistics Club.
- Bolton, T. (1894). Rhythm. *American Journal of Psychology* 6: 145-238.
- Booij, G. (1996). Cliticization as Prosodic Integration: The case of Dutch. *The Linguistic Review* 13: 219-242.
- Braun, M. (2001). Speech Mirrors Norm-tones: Absolute pitch as a normal but precognitive trait. *Acoustics Research Letters Online* 2 (3): 85-90.
- Broeckx, J.L. (1967). *Muziek en Mens I: Muziek als Verschijnsel*. Uitgeverij Metropolis: Antwerpen.
- Brown, G., Anderson, A., Shillcock, R., and G. Yule (1984). *Teaching talk: Strategies for production and assessment*. Cambridge.
- Burzio, L. (1998). Multiple Correspondence. *Lingua* 104: 79-109.
- Cairns, C. and M. Feinstein (1982). Markedness and the Theory of Syllable Structure. *Linguistic Inquiry* 13: 193-226.
- Chomsky, N. (1959). On Certain Formal Properties of Grammars. *Information and Control* 2: 137-167.
- Chomsky, N. and G. Miller (1963). Introduction to the Formal Analysis of Natural Languages. In: Luce, R.D., Bush, R.R., and E. Galanter (eds.). *Handbook of Mathematical Psychology* (pp. 269-321). John Wiley.
- Coetzee, A.W. (2004). *What it Means to be a Loser: Non-Optimal Candidates in Optimality Theory*. PhD Dissertation, University of Massachusetts, Amherst.
- Collier, G.L. and C.E. Wright (1995). Temporal Rescaling of Simple and Complex Ratios in Rhythmic Tapping. *Journal of Experimental Psychology: Human Perception and Performance* 21: 602-627.
- Cook, N.D. (2002). *Tone of Voice and Mind. The Connections between Intonation, Emotion, Cognition and Consciousness*. Amsterdam: John Benjamins.

- Cook, N.D., T.Fujisawa, and K. Takami (2004). Application of a Psycho-acoustical Model of Harmony to Speech Prosody. In: Bel, B. and I. Marlien (eds.). *Proceedings of Speech Prosody* (pp. 147-150). Nara, Japan.
- Cooper, W., and J. Eady (1986). Metrical Phonology in Speech Production. *Journal of Memory and Language* 25: 369-384.
- Couper-Kuhlen, E. (1993). *English Speech Rhythm: Form and Function in Everyday Verbal Interaction*. Amsterdam: John Benjamins.
- Crystal, D. (1991). *Dictionary of Linguistics and Phonetics*, Third Edition. Oxford: Blackwell Paperback.
- Crystal, D. (ed., 1991). *Linguistic Controversies*. London: Edward Arnold.
- Cummins, F. (1997). *Rhythmic Coordination in English Speech: An Experimental Study*. PhD Dissertation, Indiana University, Bloomington, Indiana.
- Cummins, F., and R. Port (1998). Rhythmic Constraints on Stress Timing in English. *Journal of Phonetics* 26 (2): 145-171.
- Cutler, A., Dahan, D., and W. van Donselaar (1997). Prosody in the Comprehension of Spoken Language: a Literature Review. *Language and Speech* 40 (2): 141-201.
- Das, S. (2001). *Some Aspects of the Prosodic Phonology of Tripura Bangla and Tripura Bangla English*. PhD Dissertation, CIEFL Hyderabad.
- Dasher, R. and D.L. Bolinger (1982). On Preaccentual Lengthening. *Journal of the International Phonetic Association* 12: 58-71.
- Dauer, R.M. (1983). Stress-timing and Syllable-timing Reanalysed. *Journal of Phonetics* 11: 51-62.
- Desain, P. and H. Honing (1993). Tempo Curves Considered Harmful. Time in Contemporary Musical Thought. In: Kramer, D. (ed.), *Contemporary Music Review* 7 (2) (pp. 123-138). London: Harwood Press.
- Desain, P. and H. Honing (1994). Does Expressive Timing in Music Performance Scale Proportionally with Tempo? *Psychological Research* 56: 285-292.
- Desain, P. and H. Honing (2003). The Formation of Rhythmic Categories and Metric Priming. *Perception* 32 (3): 341-365.
- Donzel, M. E. van (1997). Perception of discourse boundaries and prominence in spontaneous Dutch speech. *Working Papers Lund University* 46. Department of Linguistics and Phonetics: 5-23.
- Donzel, M. E. van (1999). *Prosodic aspects of information structure in discourse*. PhD Dissertation, University of Amsterdam.

- Dresher, B.E. and A. Lahiri (1991). The Germanic Foot: Metrical Coherence in Old English. *Linguistic Inquiry* 22: 251-286.
- Eefting, W. and T. Rietveld (1989). Just Noticeable Differences of Articulation Rate at Sentence Level. *Speech Communication* 8: 355-351.
- Eerten, L.J.A. van (2004). *Mineur en majeur in emotionele spraak, een intonatieonderzoek*. Bachelor thesis, University of Groningen.
- Eisner, J. (2000). Directional Constraint Evaluation in Optimality Theory. *Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000)*: 257-263.
- Elenbaas, N. and R. Kager (1999). Ternary rhythm and the lapse constraint. *Phonology* 16: 273-329.
- Fox, A. (2000). *Prosodic Features and Prosodic Structure*. The Phonology of Suprasegmentals. Oxford University Press.
- Fraisse, P. (1963). *The Psychology of Time*. NY: Harper and Row.
- Fraisse, P. (1982). Rhythm and Tempo. In: Deutsch, D. (ed.). *The Psychology of Music* (pp. 149-180). New York: Academic Press.
- Gilbers, D. (1987). Ritmische Structuur. *Glott* 10: 271-292.
- Gilbers, D. (1992). *Phonological Networks: a Theory of Segment Representation*. PhD Dissertation, Rijksuniversiteit Groningen.
- Gilbers, D. and H. de Hoop (1998). Conflicting Constraints: An Introduction to Optimality Theory. *Lingua* 104: 1-12.
- Gilbers, D. and W. Jansen (1996). Klemtoon en Ritme in Optimality Theory, deel 1: Hoofd-, Neven-, Samenstellings- en Woordgroepsklemtoon in het Nederlands. *TABU* 26: 53-101.
- Gilbers, D. and M. Schreuder (2000). Taal en Muziek in Optimaliteitstheorie. *TABU* 30.1-2: 1-26.
- Gilbers, D. and M. Schreuder (2002). *Language and Music in Optimality Theory*. Ms., Rutgers Optimality Archive [ROA-571].
- Gilbers, D. and M. Schreuder (in press). The Structural Resemblance of Music and Language, *Proceedings of the 7th International Congress on Musical Signification* (2001), Imatra, Finland: 530-542.
- Grabe, E. and P. Warren (1995). Stress Shift: do Speakers do it or do Listeners Hear it? In: Connell, B. and A. Arvaniti (eds.). *Phonology and phonetic evidence. Papers in Laboratory Phonology IV* (pp. 95-110). Cambridge: Cambridge University Press.
- Guéron, J. (1974). The Meter of Nursery Rhymes: an Application of the Halle-Keyser Theory of Meter. *Poetics* 12: 73-110.
- Gussenhoven, C. (1983). Stress shift in Dutch as a rhetorical device. *Linguistics* 21: 603-619.



- Gussenhoven, C. (1984). *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris.
- Gussenhoven, C. (1991). The English Rhythm Rule as an Accent Deletion Rule. *Phonology* 8: 1-35.
- Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and Phonology. *Speech Prosody 2002: Proceedings of the First International Conference on Speech Prosody*. Aix-en-Provence, ProSig and Université de Provence Laboratoire Parole et Language: 47-57.
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University.
- Gussenhoven, C. (2005). Procliticized Phonological Phrases in English: Evidence from Rhythm. *Studia Linguistica* 59: 174-193.
- Gussenhoven, C., Rietveld, T., and V. Terken (1999). *ToDI, Transcription of Dutch Intonation*, courseware. Ms., ToDI Collective. [Http://todi.let.kun.nl/](http://todi.let.kun.nl/).
- Handel S. (1993). The effect of Tempo and Tone Duration on Rhythm Discrimination. *Perception and Psychophysics* 54 (3): 370-382.
- Handel, S. (1993). The Effect of Tempo and Tone Duration on Rhythm Discrimination. *Perception and Psychophysics* 54 (3): 370-382.
- Hart, J. 't, Collier, R., and A. Cohen (1990). *A Perceptual Study of Intonation. An Experimental-Phonetic Approach to Speech Melody*. Cambridge: Cambridge University Press.
- Hauser, M.D., Chomsky, N., and W. Tecumseh Fitch (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science* 298: 1569-1579.
- Hayes, B. (1984). The Phonology of Rhythm in English. *Linguistic Inquiry* 15 (1): 33-74.
- Hayes, B. (1989). The Prosodic Hierarchy in Meter. In: Kiparsky, P. and G. Youmans (eds.). *Phonetics and Phonology, Vol. 1: Rhythm and Meter* (pp. 201-260). San Diego: Academic Press.
- Hayes, B. (1995). *Metrical Stress Theory: Principles and Case studies*. Chicago: The Chicago University Press.
- Hayes, B. and A. Kaun (1996). The Role of Phonological Phrasing in Sung and Chanted Verse. *The linguistic review* 13 (3-4): 243-304.
- Hayes, B. and M. MacEachern (1998). Quatrain Form in English Folk Verse. *Language* 74: 473-507.

- Hewitt, M. (1992). *Vertical Maximization and Metrical Theory*. PhD Dissertation, Brandeis University, Waltham, Massachusetts.
- Hint, M. (1973). *Eesti Keele Sõnafonoloogia I. Rõhusüsteemi fonoloogia ja morfonoloogia põhiprobleemid*. [Word phonology of Estonian I. The main phonological and morphophonological problems of the Estonian stress system]. Eesti NSV Teaduste Akademia, Tallinn.
- Hofstadter, D.R. (1979). *Gödel, Escher, Bach: An Eternal Golden Braid*. Vintage Books: New York.
- Honing, H. (2002). Structure and Interpretation of Rhythm and Timing. *Tijdschrift voor Muziektheorie* 7 (3): 227-232.
- Horne, M. (1990). Empirical Evidence for a Deletion Analysis of the Rhythm Rule in English. *Linguistics* 28: 959-981.
- Horne, M. (ed., 2000). *Prosody: Theory and Experiment*. Kluwer Academic Publishers: the Netherlands.
- Horne, M., Hansson, P., Bruce, G., Frid, J., and A. Jönsson (1999). Accentuation of Domain-related Information in Swedish Dialogues. *Proceedings of ESCA International Workshop on Dialogue and Prosody*. Veldhoven, The Netherlands: 71-76.
- Huss, V. (1978). English Word Stress in the Postnuclear Position. *Phonetica* 35: 86-105.
- Inkelas, S. (1989). *Prosodic Constituency in the Lexicon*. PhD Dissertation, Stanford University.
- Itô, J. and R.A. Mester (1992). *Weak Layering and Word Binariness*. Ms., Linguistic Research Center, LRC-92-09, University of California, Santa Cruz.
- Jackendoff, R. and F. Lerdahl (1980). *A Deep Parallel between Music and Language*, Indiana University Linguistic Club.
- Jakobson, R. (1932). Zur Struktur des russischen Verbums. In: Mathesio, C.G. (ed.). *Cercle Linguistique de Prague* (pp. 74-84). Prague. (Also in *Selected Writings* 2).
- Kager, R. (1993). Alternatives to the Iambic-Trochaic Law. *Natural Language and Linguistic Theory* 11: 381-432.
- Kager, R. (1994). *Ternary Rhythm in Alignment Theory*. Ms., Utrecht University. Rutgers Optimality Archive [ROA-35].
- Kager, R. and E.A.M. Visch (1988). Metrical Constituency and Rhythmic Adjustment. *Phonology* 5 (1): 21-72.
- Kawaguchi, Y. (1982). A Morphological Study of the Form of Nature. *Computer Graphics* 16 (3): 223-232.

- Keller, F. and A. Asudeh (2001). Constraints on Linguistic Coreference: Structural vs. Pragmatic Factors. In: Moore, J.D. and K. Stenning (eds.). *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 483–488). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Kelso, J.A.S. (1995). *Dynamic Patterns: The Self-Organisation of Brain and Behavior*. MIT Press.
- Kiparsky, P. (1975). Stress, Syntax, and Meter. *Language* 51 (3): 576-616.
- Kiparsky, P. (1982a). From Cyclic to Lexical Phonology. In: Hulst, H. van der and N. Smith (eds.). *The Structure of Phonological Representations, Part I*. (pp. 131-175). Dordrecht: Foris.
- Kiparsky, P. (1982b). Lexical Phonology and Morphology. In: Yang, I.S. (ed.). *Linguistics in the Morning Calm* (pp. 3-91). Linguistic Society of Korea, Hanshin, Seoul.
- Kirkpatrick, S., Gelatt, C.D. Jr., and M.P. Vecchi. (1983). Optimization by Simulated Annealing. *Science* 220 (4598): 671-680.
- Koch, H.C. (1983). *Introductory Essay on Composition*. Translated by Nancy Kovaleff Baker. New Haven, Connecticut: Yale University Press.
- Koster, J. (2003). Taal, Kunst en Biologie. *Armada, tijdschrift voor wereldliteratuur* 29/30: 85-103.
- Ladd, D.R. (1986). Intonational Phrasing: the Case for Recursive Prosodic Structure. *Yearbook of Phonology* 3: 311-340.
- Ladd, D.R. (1992) *Compound Prosodic Domains*. Occasional paper. Linguistics department, University of Edinburgh.
- Ladd, D.R. (1996) *Intonational Phonology*. Cambridge Studies in Linguistics 79, Cambridge University Press.
- Lasher, M. (1978). *A Study in the Cognitive Representation of Human Motion*. PhD Dissertation, Columbia University.
- Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.
- Lerdahl, F. and R. Jackendoff (1977). Toward a Formal Theory of Tonal Music. *Journal of music theory* 21: 111-171.
- Lerdahl, F. and R. Jackendoff (1983). *A Generative Theory of Tonal Music*. The MIT Press: Cambridge, Massachusetts, London, England.
- Liberman, M. (1975). *The Intonational System of English*. Garland Publishing Inc.: New York and London.
- Liberman, M. and A. Prince (1977). On Stress and Linguistic Rhythm. *Linguistic Inquiry* 8 (2): 249-336.

- Lindblom, B. (1978). Final lengthening in Speech and music. In: Gårding, E., Bruce, G. and R. Bannert (eds.). *Nordic Prosody: Papers from a symposium* (pp. 85-102). Lund: Lund University, Department of Linguistics.
- Linde, K. van der (2001). *Sonority Substitutions*. PhD Dissertation, Rijksuniversiteit Groningen.
- London, J. (2001). Rhythm. In: Sadie, S. (ed.). *The New Grove Dictionary of Music and Musicians* 21 Second edition. (pp. 277-309). Macmillan Publishers Limited.
- Mandelbrot, B. (1977). *Fractals: Form, Chance, and Dimension*. San Francisco: W.H. Freeman.
- McCarthy, J.J. (1986). OCP Effects: Gemination and antigemination. *Linguistic Inquiry* 17: 207-263.
- McCarthy, J.J. and A.S. Prince (1993a). *Prosodic Morphology I: Constraint Interaction and Satisfaction*. Technical Report 3, Rutgers University Center for Cognitive Science.
- McCarthy, J.J. and A.S. Prince (1993b). Generalized Alignment. In: Booij, G. and J. van Marle (eds.). *Yearbook of Morphology 1993* (pp. 79-153). Dordrecht: Kluwer.
- Milne, A.A. (1994). *Winnie de Poeh*. Dutch translation by M. Bouhuys. Van Goor: Amsterdam.
- Milne, A.A. (1996). *Het huis in het Poeh-hoekje*. Dutch translation by M. Bouhuys. Van Goor: Amsterdam.
- Morton, J., Marcus, S.M., and C.R. Frankish (1976). Perceptual Centers (P-centers). *Psychological Review* 83: 405-408.
- Nederhof, M.-J. (2000). Practical Experiments with Regular Approximation of Context-Free Languages. *Computational Linguistics* 26 (1): 17-44.
- Neijt, A. and W. Zonneveld (1982). Metrische fonologie - De Representatie van Klemtoon in Nederlandse Monomorfematische Woorden. *De nieuwe Taalgids* 75: 527-547.
- Nespor, M. and I. Vogel (1986). *Prosodic Phonology*. Dordrecht: Foris.
- Nooteboom, S.G. en A. Cohen (1995). *Spreken en Verstaan*. Van Gorcum: Assen.
- Nouveau, D. (1994). *Language Acquisition, Metrical Theory, and Optimality: A Study of Dutch Word Stress*. PhD Dissertation, Rijksuniversiteit Utrecht.
- Noys, B. (1995). Into the 'Jungle'. *Popular Music* 14/3. Cambridge University Press: 321-332.

- Oehrle, R. (1989). Temporal Structures in Verse Design. In: P.Kiparsky and G. Youmans (eds.). *Rhythm and Meter* (pp. 87-119). San Diego: Academic Press.
- Oostendorp, M. van (1995). *Vowel Quality and Syllabic Projection*. PhD Dissertation, Katholieke Universiteit Brabant.
- Ouden, D.B. den (2004). Multi-level OT: An Argument from Speech Pathology. *Linguistics in the Netherlands* 21 (1): 146-157.
- Ouden, H. den (2004). *Prosodic Realizations of Text Structure*. PhD Dissertation, University of Tilburg.
- Palmer, C. (1989). Mapping Musical Thought to Musical Performance. *Journal of Experimental Psychology: Human Perception and Performance* 15: 331-346.
- Palmer, C. (1997). Music Performance. *Annual Review of Psychology* 48: 115-138.
- Patel, A.D. (1998). Syntactic Processing in Language and Music: Different Cognitive Operations, Similar Neural Resources? *Music Perception* 16 (1): 27-42.
- Patel, A.D. (2003). Language, Music, Syntax and the Brain. *Nature Neuroscience* 6 (7), *Special Issue: Focus on Music*: 674-681.
- Patel, A.D. and J.R. Daniele (2003). An Empirical Comparison of Rhythm in Language and Music. *Cognition* 87: B35-B45.
- Patel, A.D., Gibson, E., Ratner, J., Besson, M., and P.J. Holcomb (1996). Processing Grammatical Relations in Music and Language: An Event-related Potential (ERP) Study. *Proceedings of the Fourth International Conference on Music Perception and Cognition*. Montreal: McGill University, Faculty of Music: 337-342.
- Patel, A.D., Löfqvist, A., and W. Naito (1999). The Acoustics and Kinematics of Regularly-timed Speech: A Database and Method for the Study of the P-center Problem. *Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco*. 1: 405-408.
- Patel, A.D. and I. Peretz (1997). Is Music Autonomous from Language? A Neuropsychological Appraisal. In: Deliège, I. and J. Sloboda (eds.). *Perception and Cognition of Music* (pp. 191-215). London: Erlbaum Psychology Press.
- Patel, A.D., Peretz, I., Tramo, M., and R. Labrecque (1998a). Processing Prosodic and Musical Patterns: A Neuropsychological Investigation. *Brain and Language* 61: 123-144.
- Patel, A.D., Gibson, E., Ratner, J., Besson, M., and P.J. Holcomb (1998b). Articles - Processing Syntactic Relations in Language

- and Music: An Event-Related Potential Study. *Journal of cognitive neuroscience* 10: 717-733.
- Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. PhD Dissertation, MIT.
- Pike, K.L. (1945). *The Intonation of American English*. Ann Arbor, Michigan: University of Michigan Press.
- Platel, H., Price, C., Baron, J.-C., Wise, R., Lambert, J., Frackowiak, R.S.J., Lechevalier, B., and F. Eustache (1997). The Structural Components of Music Perception. A Functional Anatomical Study. *Brain* 120: 229-243.
- Pompino-Marshall, B. (1991) The Syllable as a Prosodic Unit and the So-called P-centre Effect. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München* 29: 65-123.
- Port, R.F., Tajima, K., and F. Cummins (1998). Speech and Rhythmic Behavior. In: Savelsburgh, G.J.P., Maas, H. van der, and P.C.L. van Geert (eds.). *Non-linear Developmental Processes* (pp. 53-78). Elsevier: Amsterdam.
- Povel D.J. (1981). Internal Representation of Simple Temporal Patterns. *Journal of Experimental Psychology, Human Perception and Performance* 7 (1): 3-18.
- Prince, A. (1983). Relating to the Grid. *Linguistic Inquiry* 14: 19-100.
- Prince, A. and P. Smolensky (1993, a.k.a. 2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Ms., Rutgers Optimality Archive. Rutgers Optimality Archive [ROA-537]. (2004: Blackwell Publishing.)
- Quené, H. (2003). *Over het Perceptieve Belang van Ritme en Metrum*. Paper presented at the Dag van de Fonetiek, December 2003, Universiteit Utrecht.
- Quené, H., and R.F. Port (2003). Rhythmical Factors in Stress Shift. In: M. Andronis, E. Debenport, A. Pycha, and K. Yoshimura, (eds.). *CLS 38: Papers from the 38th Meeting of the Chicago Linguistic Society, vol. 1: Main Session*. Chicago Linguistic Society, Chicago.
- Quené, H. and R.F. Port (2005). Effects of Timing Regularity and Metrical Expectancy on Spoken-Word Perception. *Phonetica* 62: 1-13.
- Raffman, D.L. (1994). Language, Music, and Mind. *Journal of Linguistics* 30 (2): 575-578.
- Ramus, F. (2002). Acoustic Correlates of Linguistic Rhythm: Perspectives. *Proceedings of Speech Prosody 2002, Aix-en-*

- Provence*. Aix-en-Provence: Laboratoire Parole et Langage: 115-120.
- Ramus, F., Dupoux, E., and J. Mehler (2003). The Psychological Reality of Rhythm Classes: Perceptual Studies. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona: 337-342.
- Ramus, F., Nespor, M., and J. Mehler (1999). Correlates of Linguistic Rhythm in the Speech Signal. *Cognition* 73 (3): 265-292.
- Randel, D.M. (ed., 1986). *The New Harvard Dictionary of Music*. The Belknap Press of Harvard University Press: Cambridge, Massachusetts, London, England.
- Reeves C.R. (ed., 1995). *Modern Heuristic Techniques for Combinatorial Problems*. McGraw-Hill, London, etc.
- Reicha, A. (1814). *Traité de Mélodie*. Paris.
- Repp, B.H. (1990). Patterns of Expressive Timing in Performances of Beethoven Minuet by Nineteen Famous Pianists. *Journal of the Acoustical Society of America* 88: 622-641.
- Repp, B.H. (1994). Relational Invariance of Expressive Microstructure Across Global Tempo Changes in Music Performance: An Exploratory Study. *Psychological Research* 56: 269-284.
- Repp, B.H. (1995). Quantitative Effects of Global Tempo on Expressive Timing in Music Performance: some Perceptual Evidence. *Music Perception* 13: 39-57.
- Repp, B.H. (1998). A Microcosm of Musical Expression. I. Quantitative Analysis of Pianists' Timing in the Initial Measures of Chopin's Etude in E major. *Journal of the Acoustical Society of America* 104 (2): 1085-1100.
- Repp, B.H. (2000). Subliminal Temporal Discrimination Revealed in Sensorimotor Coordination. In: Desain, P. and L. Windsor (eds.). *Rhythm Perception and Production* (pp. 129-142). Lisse, The Netherlands: Swets and Zeitlinger.
- Repp, B.H. (2002). The Embodiment of Musical Structure: Effects of Musical Context on Sensorimotor Synchronization with Complex Timing Patterns. In: Prinz, W. and B. Hommel (eds.). *Common mechanisms in perception and action: Attention and Performance XIX* (pp. 245-265). Oxford, U.K.: Oxford University Press.
- Repp, B.H., Windsor, W.L., and P. Desain (2002). Effects of Tempo on the Timing of Simple Musical Rhythms. *Music Perception* 19 (40): 565-593.

- Reynolds, S. (1999). *Geration Ecstasy: into the World of Techno and Rave Culture*. Boston: Little, Brown, and Company.
- Riemann, H. (1902). *Grosse Kompositionslehre. Vol. I: Der homophone Satz*. Berlin: W. Spermann.
- Rietveld, A.C.M. and V.J. van Heuven (1997). *Algemene Fonetiek*. Dick Coutinho, Bussum.
- Roach, P. (1982). On the Discrimination between “Stress-timed” and “Syllable-timed” Languages. In: Crystal, D. (ed.): *Linguistic controversies* (pp. 73-79). London: Edward Arnold.
- Rothstein, W.N. (1989). *Phrase Rhythm in Tonal Music*. New York, N.Y: Schirmer Books, London: Collier Macmillan.
- Sachs, C. (1953). *Rhythm and Tempo: a Study in Music History*. New York, N.Y: Norton.
- Samek-Lodovici, V. and A. Prince (1999). *Optima*. RuCCS-TR-57. Rutgers Center for Cognitive Science. New Brunswick: Rutgers University. Also available as Rutgers Optimality Archive [ROA-363].
- Schreuder, M. (1999). *Taal en Muziek, de Structurele Tweeling van ons Cognitieve Systeem*. Doctoral thesis, Groningen.
- Schreuder, M., and D. Gilbers (2004a). Recursive Patterns in Phonological Phrases. In: Bel, B. and I. Marlien (eds.). *Proceedings of Speech Prosody 2004* (pp. 341-344). Nara, Japan.
- Schreuder, M. and D. Gilbers (2004b). The Influence of Speech Rate on Rhythm Patterns. In: Gilbers, D., Schreuder, M., and N. Knevel (eds.). *On the Boundaries of Phonology and Phonetics*. (pp. 183-202). Rijksuniversiteit Groningen.
- Schreuder, M., Eerten, L. van, and D. Gilbers (2006). Music as a Method of Identifying Emotional Speech. In: Devillers, L. and J.C. Martin (eds.). *Workshop LREC 2006 on Corpora for Research on Emotion and Affect* (pp. 55-59). Genua, Italy. LIMSI-CNRS, FR, R. Cowie, E. Douglas-Cowie, QUB, UK, A. Batliner, Erlangen University, Germany.
- Schreuder, M. and D. Gilbers (in press). Phrasing in Language and Music. *Proceedings of the 7th International Congress on Musical Signification* (2001), Imatra, Finland: 543-553.
- Schreuder, M., and D. Gilbers (to appear). Restructuring the Melodic Content of Feet. *Proceedings of the 9th International Phonology Meeting 2002*, Vienna, Austria.
- Selkirk, E. (1978). On Prosodic Structure in Relation to Syntactic Structure. In: T. Fretheim (ed.). *Nordic Prosody 2* (pp. 111-140). TAPIR: Trondheim, Norway.



- Selkirk, E. (1980). The Role of Prosodic Categories in English Word Stress. *Linguistic Inquiry* 11: 563-605.
- Selkirk, E. (1984). *Phonology and Syntax: The Relation Between Sound and Structure*. Cambridge, Massachusetts: MIT Press.
- Selkirk, E. (1995a). Sentence Prosody: Intonation, Stress, and Phrasing. In: J. Goldsmith (ed.). *The handbook of Phonological Theory* (pp. 550-569). Blackwell Reference.
- Selkirk, E. (1995b). The Prosodic Structure of Function Words. In: Beckman, J. et al. (eds.). *Papers in Optimality Theory*. University of Massachusetts Occasional Papers in Linguistics 18 (pp. 439-469). Amherst, MA: Graduate Linguistic Student Association.
- Selkirk, E. (2000). The interaction of constraints on prosodic phrasing. In: Horne, M. (ed.): 231-261.
- Shattuck-Hufnagel, S. (2000). Phrase-level Phonology in Speech Production Planning: Evidence for the Role of Prosodic Structure. In: Horne, M. (ed.): 201-231.
- Shattuck-Hufnagel, S., Ostendorf, M., and K. Ross (1994). Stress Shift and Early Pitch Accent Placement in Lexical Items in American English. *Journal of Phonetics* 22: 357-388.
- Shattuck-Hufnagel, S. and A.C. Turk (1996). A Prosody Tutorial for Investigators of Auditory Sentence Processing, *Journal of Psycholinguistic Research* 25 (2): 193-247.
- Sluijter, A. (1995). *Phonetic Correlates of Stress and Accent*. PhD Dissertation, HIL Dissertations 15. Leiden University.
- Sluijter, A. and V. van Heuven (1996). Spectral Balance as an Acoustic Correlate of Linguistic Stress. *Journal of the Acoustical Society of America* 100 (4): 2471-2485.
- Smit, B. de and H.W. Lenstra Jr. (2003). The Mathematical Structure of Escher's Print Gallery. *Notices of the AMS* 50 (4): 446-451.
- Smith, A.R. (1984). Plants, Fractals, and Formal Languages. *Computer Graphics* 18 (3): 1-10.
- Solomon, L. (1998). *The Fractal Nature of Music*. Tucson, Arizona.
- Stowe, L., Haverkort, M. and F. Zwarts (2005). Rethinking the Neurological Basis of Language. *Lingua* 115: 997-1042.
- Swerts, M. (1994). *Prosodic features of discourse units*. PhD Dissertation, Eindhoven University.
- Terken, J. and D. Hermes (2000). The Perception of Prosodic Prominence. Horne, M. (ed.): 89-127.
- Tesar, B. and P. Smolensky (1998). Learnability in Optimality Theory. *Linguistic Inquiry* 29: 229-268.

- Todd, N.P.M. (1985). A Model of Expressive Timing in Tonal Music. *Music Perception* 3: 33–58.
- Todd, N.P.M. (1989). A Computational Model of Rubato. *Contemporary Music Review* 3:69–88.
- Trubetskoy, N. (1939). *Grundzüge der Phonologie*. = *Travaux du Cercle Linguistique de Prague* 7. Repr. (1968). Göttingen: Vandenhoeck and Ruprecht.
- Truckenbrodt, H. (1999). On the Relation between Syntactic Phrases and Phonological Phrases. *Linguistic Inquiry* 30: 219-255.
- Vigário, M. (1999). On the Prosodic Status of Stressless Function Words in European Portuguese. In: Hall, T. and U. Kleinhenz (eds.). *Studies on the Phonological Word*. (pp. 255-295). John Benjamins Publishing Company: Amsterdam/Philadelphia.
- Visch, E.A.M. (1989). *A Metrical Theory of Rhythmic Stress Phenomena*. PhD Dissertation, Rijksuniversiteit Utrecht.
- Vogel, I., Bunnell, T., and S. Hoskins (1995). The Phonology and Phonetics of the Rhythm Rule. In: Connell, B. and A. Arvaniti, (eds.). *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV* (pp. 111-127). Cambridge. Cambridge University Press.
- Wallin, N.L. (1991). *Biomusicology: Neurophysiological, Neuropsychological and Evolutionary Perspectives on the Origins and Purposes of Music*. Stuyvesant, New York: Pendragon Press.
- Wenk, B.J. (1987). Just in time: on speech rhythms in music. *Linguistics* 25: 969–981.
- Woodrow, H. (1951). Time perception. In: Stevens, S. S. (ed.). *Handbook of Experimental Psychology* (pp. 1224-1236). New York: Wiley.
- Zec, D. and S. Inkelas (1991). The Place of Clitics in the Prosodic Hierarchy. *Proceedings of the 10<sup>th</sup> West Coast Conference on Formal Linguistics*: 505-519.
- Zonneveld, R.M. van (1983). *Affix Grammatica: Een Onderzoek naar woordvorming in het Nederlands*. PhD Dissertation, Rijksuniversiteit Groningen.
- Zonneveld, R.M. van (1988). Two Level Phonology: Structural Stability and Segmental Variation in Dutch Child Language. In: Besien, F. van (ed.). *First Language Acquisition, ABLA papers* 12 (pp. 129-162). University of Antwerpen.

## Samenvatting

Iedereen is wel bekend met het fenomeen van de tikkende klok. Een klok zegt *tik tak, tik tak*. Waarom kennen wij de twee tikken van de klok twee verschillende klinkers toe? Klinken de twee tikken dan verschillend? Van sommige klokken wel, maar zeker niet van allemaal. De twee tikken zijn vaak nauwelijks van elkaar te onderscheiden. Het is onze eigen fantasie die een *tik* aan de eerste tik toekent en een *tak* aan de tweede. Het feit dat iedereen dit doet zegt iets over de manier waarop onze cognitie werkt.

Hoe werkt die dan? Onze cognitie wil graag structuur horen om alles om ons heen goed te begrijpen, in hap-klare brokken. En het wil de samenhang tussen de elementen kennen. Die samenhang wordt herkend door alles onder te brengen in hiërarchieën, met verschillende niveaus en op elk niveau is altijd één element het meest belangrijke, het hoofdelement. Bij het tikken van de klok besluiten we dat de twee tikken een groep van twee vormen en het eerste element van deze groep van twee tikken is het belangrijkste. We schrijven het een /t/-klank toe, een intrinsiek hogere vocal dan de /d/-klank. We denken het ook waar te nemen als luider, langer in duur en hoger van toonhoogte.

Door alle elementen van een geluidsobject – of zelfs van visuele objecten of bewegingen – onder te brengen in een hiërarchie van belangrijke en minder belangrijke elementen wordt het interpreteren gemakkelijker gemaakt. Dit proefschrift gaat over taal en muziek, die beide vallen onder cognitief menselijk gedrag en beide te maken hebben met geluid. In het eerste hoofdstuk baseren we ons voor een belangrijk deel op het boek *Een Generatieve Theorie van Tonale Muziek* van Lerdahl en Jackendoff. In de jaren zeventig volgden de taalkundige Ray Jackendoff en de musicoloog Fred Lerdahl een seminar, gegeven door Leonard Bernstein over de structuur van muziek. Het viel hen op dat er grote overeenkomsten zijn tussen de manier waarop enerzijds taalkundigen en anderzijds musicologen hun object van onderzoek structureren. Om muziek of taal te kunnen vatten luisteren we niet naar een stroom opeenvolgende gelijkwaardige klanken, maar proberen we vaak onbewust verbanden te leggen tussen de verschillende klanken en proberen we tevens vast

te stellen welke klanken essentieel zijn en welke minder belangrijk. Als een luisteraar niet vaststelt wat de belangrijke onderdelen zijn van het geheel, is hij niet in staat de structuur te doorgronden en haakt hij af.

Jackendoff en Lerdahl gingen ervan uit dat wij muziek en taal op dezelfde wijze in ons brein verwerken. Dit inzicht was voor hen aanleiding om een theorie voor tonale muziek voor te stellen waarmee ze muzikale intuïtie willen beschrijven. Met name inzichten uit de toenmalige leidende stroming in de fonologie hebben ertoe geleid dat in partituren wordt aangegeven wat de belangrijke (hoofden) en wat de minder belangrijke onderdelen (afhankelijken) zijn in het muziekstuk. Op deze manier brengen Lerdahl en Jackendoff een synthese tot stand van taalkunde en muziektheorie. Ze maakten gebruik van boom- en gridstructuren, veelgebruikte gereedschappen in de syntaxis en de fonologie. Deze representatiemethoden gebruikten ze om zichtbaar te maken hoe muziek hiërarchisch is opgebouwd.

De methodologie die Lerdahl en Jackendoff in 1983 introduceerden is gebaseerd op preferentieregels, om te beschrijven hoe we komen tot een ideale interpretatie van een muziekstuk. Deze methode van preferentieregels werd tien jaar later opgevolgd door Prince en Smolensky (1983) met hun inmiddels voor de fonologie toonaangevende Optimaliteitstheorie. De schendbare OT-constraints laten een opmerkelijke overeenkomst zien met de preferentieregels voor muziek. De constraints of preferentieregels bepalen wat grammaticaal is in taal en welke interpretatie optimaal is in muziek en in beide theorieën zijn de preferentieregels ‘zacht’ en potentieel conflicterend, wat de theorieën hun kracht geeft. In Hoofdstuk 1 wijzen we op deze gelijkenis en vergelijken we een paar muzikale preferentieregels met OT-constraints.

Er zijn vele manieren die tot de beslissing leiden welke elementen de belangrijkste zijn van een geluidsobject als taal of muziek, en welke elementen samen domeinen vormen als lettergrepen, voeten of frases. In taal en muziek worden deze domeinen bepaald op basis van samenhang in betekenis, structuur, of vorm, of op basis van afstand tot of verschil met andere elementen. De groeperingen op basis van betekenis in taal (semantiek) kunnen verschillen van de groeperingen op basis de fonologische structuur, die op hun beurt weer kunnen

afwijken van die op basis van de syntactische structuur. Intonatie, ritme, pauzes, etc. voegen hun eigen groeperingsverschijnselen toe aan waargenomen teksten. In muziek spelen vergelijkbare invloeden een structurerende rol: melodielijnen, ritme, rusten, harmonie, etc. De cognitie van de luisteraar hecht belang aan sommige elementen in de muziek of het spraakgeluid, terwijl andere elementen gezien kunnen worden als versiering. Deze keuze wordt gemaakt op basis van alle verschillende cues voor structuur en groepering en de cognitie heeft zijn eigen voorkeursregels ('constraints') voor al deze aanwijzingen. De rangschikking van constraints of preferentieregels bepaalt welke cues het belangrijkste zijn.

Op basis van de overeenkomsten tussen taal en muziek stellen we voor dat de fonoloog kan profiteren van de kennis van de muzikoloog. Muziektheorie kan misschien helpen om taalkundige kwesties op te lossen die moeilijk op te lossen zijn met de taalkundige theorie alleen. Drie van dat soort kwesties onderzoeken we experimenteel.

Hoofdstuk 2 is een introductie van de achtergrondtheorie over ritme. Ritme kan gezien worden als het meest duidelijke gedeelde kenmerk van taal en muziek. Het hoofdstuk geeft een overzicht van een aantal theoretische onderwerpen die te maken hebben met ritme. In de eerste paragraaf laten we zien dat ritme overal aanwezig is in de wereld om ons heen. We laten zien dat ritme in taal en muziek bestaat uit regelmatige of onregelmatige geluidspatronen, die gegroepeerd zijn in structuren waarin accenten de prominente elementen markeren.

In de Hoofdstukken 3, 4 en 5 beschrijven we de resultaten van drie prosodie-experimenten. De eerste kwestie die we onderzoeken, in Hoofdstuk 3, is de vraag of de invloed van een hogere spreek snelheid leidt tot aanpassing van de fonologische structuur of slechts tot fonetische compressie, of misschien alleen tot een andere waarneming door de luisteraar. We onderzoeken deze vraag vanuit verschillende uitgangspunten. We laten zien dat ritme in snelle spraak vaker als hergestructureerd wordt waargenomen dan in normaal spreektempo. Akoestische metingen wijzen uit dat deze waarneming niet gebaseerd is op kenmerken in het geluidssignaal,

maar op timing. Luisteraars horen op regelmatige afstanden een accent, ook al is het er niet. Mensen blijken te horen wat ze willen horen.

De tweede kwestie, in Hoofdstuk 4, is dat er in de taalkunde een scheve verhouding lijkt te bestaan tussen syntactische en fonologische structuur. Syntactische constituenten laten recursie zien, terwijl wordt aangenomen dat deze recursie geen rol speelt in de fonologie. In muziek komen recursieve, ingebedde frasestructuren ook voor. Waarom zou taalkundige prosodie zich dan verschillend gedragen van zowel syntaxis als muziek? Sterker nog, recursie komt voor in alle soorten van kunst en zelfs in de natuur. Paragraaf 2 van Hoofdstuk 4 geeft een aantal voorbeelden van verschillende soorten recursie; in de rest van Hoofdstuk 4 zoeken we naar evidentie voor het idee dat fonologie ook recursieve structuren kent. We hebben een experiment uitgevoerd waarin we een vorm van frasestructuur onderzochten. Daarin keken we of grensmarkeringsprocessen, zoals accentverschuiving, recursief toegepast kunnen worden op fonologische frases die zijn ingebed in grotere fonologische frases en we laten op basis van de resultaten zien dat recursie in fonologische frases moet worden toegelaten tot de prosodische hiërarchie. Opvallend was dat de verschoven accenten niet te meten waren in de lettergrepen waarin ze werden waargenomen. Er vond wel een verandering plaats in de lettergrepen waar de hoofdklemtoon van het woord oorspronkelijk lag. Daar bleken de waarden van de accentcorrelaten verlaagd te zijn. In andere woorden, mensen nemen een verandering waar op een andere plek dan waar de verandering werkelijk plaatsvindt.

De derde kwestie, in Hoofdstuk 5, gaat over extra-linguistische intonatie. Het draait daar om de vraag of verschillen in emotionele spraak door verschillende modaliteiten wordt gekenmerkt. In muziek wordt het verschil tussen sombere en vrolijke muziek vaak gekarakteriseerd door het verschil tussen een mineur- en een majeurtoonsoort. In dit onderzoek analyseerden we geclusterde frequentiepieken en partituren van verhaaltjes waarin de vrolijke Teigetje en de sombere Iejaar opgevoerd worden als sprekende personages. De resultaten laten zien dat in de gevallen waarin we

tertsintervallen tegenkomen de Teigetjeverhalen uitsluitend worden gekenmerkt door majeuremodaliteit (grote terts), terwijl de Iejoorverhalen de mineurmodaliteit (kleine terts) tentoonspreiden. Het lijkt er dus op, op basis van dit kleinschalige experiment, dat de uitdrukking van gemoedstoestand van emotionele spraak sterk overeenkomt met muzikale modaliteit. We kunnen hier geen verstrekkende conclusies aan verbinden, maar in ieder geval lijkt dit een interessante methodologie om emotionele intonatie te onderzoeken.

Deze prosodische vraagstukken hebben met elkaar gemeen dat de muziektheorie kan helpen met de oplossing. Alle drie onderwerpen, die gaan over ritme, frasestructuur en intonatie of melodie, zijn ook de basiselementen in de muziektheorie. Ze kunnen worden gezien als de bouwstenen van een groter geheel, zowel in taal als in muziek, die hiërarchische structuren bouwen van geluid. Zonder structuur kunnen we het eenvoudig niet begrijpen. Net als in het simpele voorbeeld van het tikken van een klok horen we structuur in elk deel en we verbinden dit aan de kenmerken van het beluisterde signaal, net zo lang tot we het complete stuk muziek of de hele gesproken tekst hebben gereconstrueerd.

Voor de experimenten hebben we steeds alle waarnemingen geprobeerd empirisch te ondersteunen met akoestische metingen. De laatste tientallen jaren doen veel fonologen dat en de opkomst van de zogenaamde Laboratory Phonology is een product van deze ambitie. Het idee daarachter is dat alle waargenomen fonologische fenomenen veroorzaakt moeten zijn door kenmerken in het akoestische signaal. Een aantal fonologische verschijnselen is zelfs alleen te verklaren met de hulp van akoestische informatie. In de experimenten die wij in Hoofdstukken 3 en 4 beschrijven trekken we echter de conclusie dat niet alles wat we horen in het geluidssignaal gevonden kan worden. In de psychologie, psycholinguïstiek en ook in de musicologie worden veel voorbeelden gegeven die aantonen dat auditieve illusies van het brein van de luisteraar een belangrijke rol spelen bij de waarneming. De luisteraar verwacht een belangrijke lettergreep op een bepaald punt in de tijd, op de beat, en hoort daarom een accent op dat punt, of het ook werkelijk in het signaal

aanwezig is of niet. Er wordt, net als in muziek, meer regelmaat waargenomen dan er in werkelijkheid is. Die strategie werkt anticiperend en speelt mogelijk een belangrijke rol om de communicatie te vergemakkelijken. Deze psychologische invloed krijgt in de fonologie weinig aandacht. Meten is niet altijd weten.



## Groningen Dissertations in Linguistics (GRODIL)

---

1. Henriëtte de Swart (1991). *Adverbs of Quantification: A Generalized Quantifier Approach.*
2. Eric Hoekstra (1991). *Licensing Conditions on Phrase Structure.*
3. Dicky Gilbers (1992). *Phonological Networks. A Theory of Segment Representation.*
4. Helen de Hoop (1992). *Case Configuration and Noun Phrase Interpretation.*
5. Gosse Bouma (1993). *Nonmonotonicity and Categorical Unification Grammar.*
6. Peter I. Blok (1993). *The Interpretation of Focus.*
7. Roelien Bastiaanse (1993). *Studies in Aphasia.*
8. Bert Bos (1993). *Rapid User Interface Development with the Script Language Gist.*
9. Wim Kosmeijer (1993). *Barriers and Licensing.*
10. Jan-Wouter Zwart (1993). *Dutch Syntax: A Minimalist Approach.*
11. Mark Kas (1993). *Essays on Boolean Functions and Negative Polarity.*
12. Ton van der Wouden (1994). *Negative Contexts.*
13. Joop Houtman (1994). *Coordination and Constituency: A Study in Categorical Grammar.*
14. Petra Hendriks (1995). *Comparatives and Categorical Grammar.*
15. Maarten de Wind (1995). *Inversion in French.*
16. Jelly Julia de Jong (1996). *The Case of Bound Pronouns in Peripheral Romance.*
17. Sjoukje van der Wal (1996). *Negative Polarity Items and Negation: Tandem Acquisition.*

18. Anastasia Giannakidou (1997). *The Landscape of Polarity Items*.
19. Karen Lattewitz (1997). *Adjacency in Dutch and German*.
20. Edith Kaan (1997). *Processing Subject-Object Ambiguities in Dutch*.
21. Henny Klein (1997). *Adverbs of Degree in Dutch*.
22. Leonie Bosveld-de Smet (1998). *On Mass and Plural Quantification: The case of French 'des'/'du'-NPs*.
23. Rita Landeweerd (1998). *Discourse semantics of perspective and temporal structure*.
24. Mettina Veenstra (1998). *Formalizing the Minimalist Program*.
25. Roel Jonkers (1998). *Comprehension and Production of Verbs in aphasic Speakers*.
26. Erik F. Tjong Kim Sang (1998). *Machine Learning of Phonotactics*.
27. Paulien Rijkhoek (1998). *On Degree Phrases and Result Clauses*.
28. Jan de Jong (1999). *Specific Language Impairment in Dutch: Inflectional Morphology and Argument Structure*.
29. H. Wee (1999). *Definite Focus*.
30. Eun-Hee Lee (2000). *Dynamic and Stative Information in Temporal Reasoning: Korean tense and aspect in discourse*.
31. Ivilin P. Stoianov (2001). *Connectionist Lexical Processing*.
32. Klarien van der Linde (2001). *Sonority substitutions*.
33. Monique Lamers (2001). *Sentence processing: using syntactic, semantic, and thematic information*.
34. Shalom Zuckerman (2001). *The Acquisition of "Optional" Movement*.
35. Rob Koeling (2001). *Dialogue-Based Disambiguation: Using Dialogue Status to Improve Speech Understanding*.
36. Esther Ruigendijk (2002). *Case assignment in Agrammatism: a cross-linguistic study*.

37. Tony Mullen (2002). *An Investigation into Compositional Features and Feature Merging for Maximum Entropy-Based Parse Selection.*
38. Nanette Bienfait (2002). *Grammatica-onderwijs aan allochtone jongeren.*
39. Dirk-Bart den Ouden (2002). *Phonology in Aphasia: Syllables and segments in level-specific deficits.*
40. Rienk Withaar (2002). *The Role of the Phonological Loop in Sentence Comprehension.*
41. Kim Sauter (2002). *Transfer and Access to Universal Grammar in Adult Second Language Acquisition.*
42. Laura Sabourin (2003). *Grammatical Gender and Second Language Processing: An ERP Study.*
43. Hein van Schie (2003). *Visual Semantics.*
44. Lilia Schürcks-Grozeva (2003). *Binding and Bulgarian.*
45. Stasinos Konstantopoulos (2003). *Using ILP to Learn Local Linguistic Structures.*
46. Wilbert Heeringa (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance.*
47. Wouter Jansen (2004). *Laryngeal Contrast and Phonetic Voicing: A Laboratory Phonology.*
48. Judith Rispens (2004). *Syntactic and phonological processing in developmental dyslexia.*
49. Danielle Bougaïré (2004). *L'approche communicative des campagnes de sensibilisation en santé publique au Burkina Faso: Les cas de la planification familiale, du sida et de l'excision.*
50. Tanja Gaustad (2004). *Linguistic Knowledge and Word Sense Disambiguation.*

51. Susanne Schoof (2004). *An HPSG Account of Nonfinite Verbal Complements in Latin*.
52. M. Begoña Villada Moirón (2005). *Data-driven identification of fixed expressions and their modifiability*.
53. Robbert Prins (2005). *Finite-State Pre-Processing for Natural Language Analysis*.
54. Leonoor van der Beek (2005) *Topics in Corpus-Based Dutch Syntax*
55. Keiko Yoshioka (2005). *Linguistic and gestural introduction and tracking of referents in L1 and L2 discourse*.
56. Sible Andringa (2005) *Form-focused instruction and the development of second language proficiency*.
57. Joanneke Prenger (2005) *Taal telt! Een onderzoek naar de rol van taalvaardigheid en tekstbegrip in het realistisch wiskundeonderwijs*.
58. Neslihan Kansu-Yetkiner (2006) *Blood, Shame and Fear: Self-Presentation Strategies of Turkish Women's Talk about their Health and Sexuality*.
59. Mónika Z. Zempléni (2006) *Functional imaging of the hemispheric contribution to language processing*.
60. Maartje Schreuder (2006) *Prosodic Processes in Language and Music*.

GRODIL

Secretary of the Department of General Linguistics

P.O. Box 716

9700 AS Groningen

The Netherlands