

Chapter 3

The Influence of Speech Rate on the Perception of Rhythm Patterns

3.1. Introduction⁹

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

⁹ 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 *stúdiëtòelàge* (S w s w w) ‘study grant’, we assume that this compound can be realized as *stúdiëtòelàge* (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 *pèrfèctioníst* (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 Data¹⁰

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’

¹⁰ Some examples from the experiment can be downloaded as mp3-files from <http://home.planet.nl/~schre537/sounds.htm> or 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

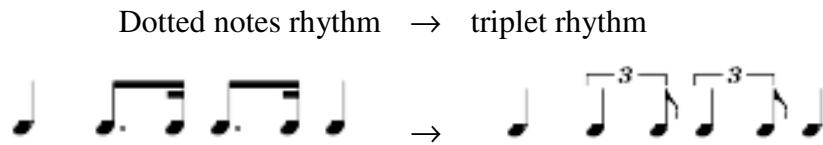


Figure 18 Dotted notes rhythm in the second movement of Beethoven's 5th

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 (* $\Sigma\Sigma$) (Kager 1994), and OUTPUT - OUTPUT CORRESPONDENCE constraints (*cf.* Burzio 1998) within the framework of OT.¹¹ 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

¹¹ Elenbaas and Kager 1999 and Das 2001 replace * $\Sigma\Sigma$ with *LAPSE, which interacts with the constraints ALL-FT-R and Parse- σ to account for ternary rhythm. To reach the same effect as * $\Sigma\Sigma$, they had to extend the definition of *LAPSE and to assume a gradient constraint ALL-FT-R. We choose to fall back on * $\Sigma\Sigma$, 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 17.¹²

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

¹² 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 Bangla¹³: 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

¹³ 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)).

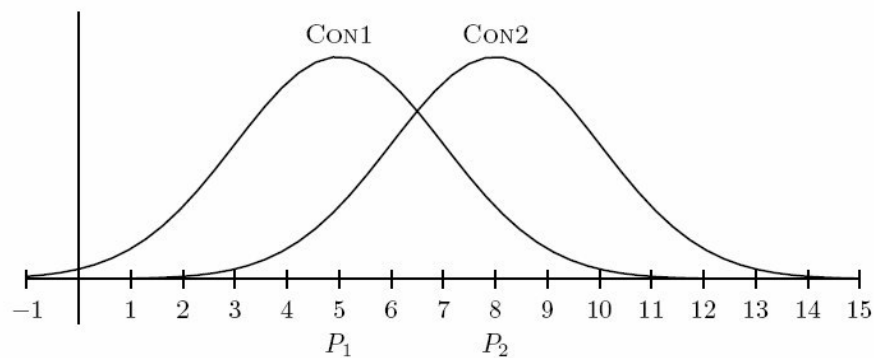
Anttila (1997) and Anttila and Cho's (1998) proposal uses a stratified hierarchy in which a subset of the constraints may be unranked in relation to each other. In their proposal the correspondence constraints and the markedness constraints should be placed on different randomly ranked strata, in order 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 * $\Sigma\Sigma$ and OUTPUT - OUTPUT CORRESPONDENCE unordered with respect to each other. Then, both outputs – (*stúdie*)(*tòelàge*) 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 * Σ . This means that the chance of selecting points for which the ranking O-O CORR >> * Σ 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 Bíró (2005 / to appear). Bíró'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' (Bíró to appear, Bíró, 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*.

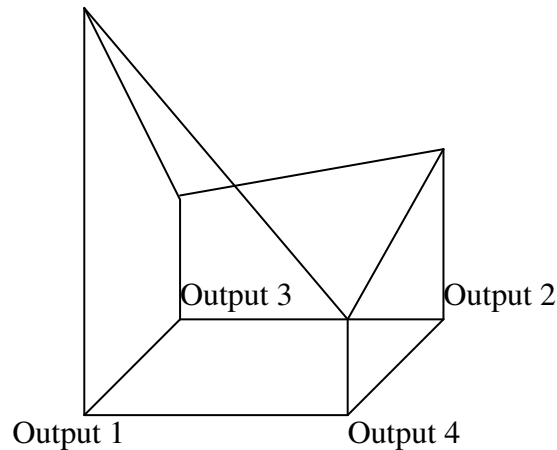
Let us now change the rule, and introduce Simulated Annealing: it is now also permissible to move downhill. Before each step, one can randomly choose a direction, horizontal, or uphill, or downhill. If the neighbor chosen is higher or equally high, the random walker certainly moves there, whereas if 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.¹⁴ 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.

¹⁴ 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.

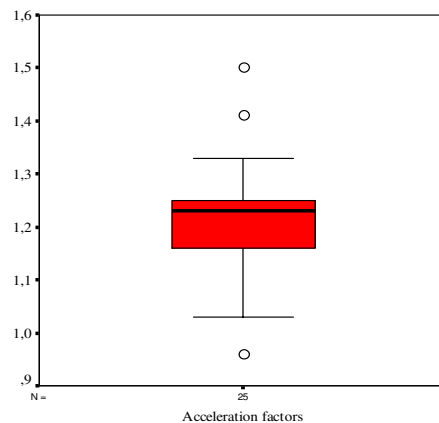
Finally, we considered rhythmic timing. The idea is that the beats in speech are separated from each other at an approximately equal distance independent of the speech rate. In other words, a speaker more or less follows an imaginary metronome. If he/she speaks faster, more melodic content will be placed 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

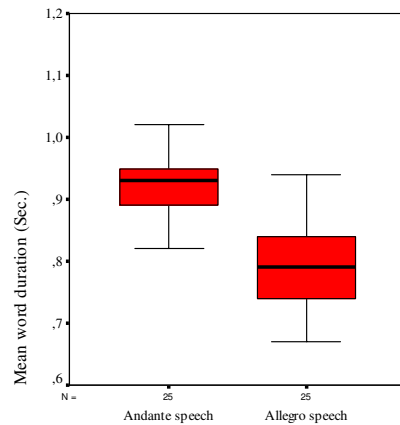
As for the pilot experiment, before looking into the auditory results, we first investigated whether the quiz design was successful: did the subjects speak faster in the quiz task than when we asked them to speak at a moderate speaking rate? We calculated the acceleration factors by dividing the mean 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



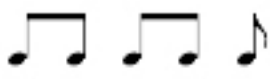



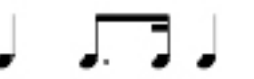

From these data we can conclude that the quiz design was successful in eliciting fast speech.

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>

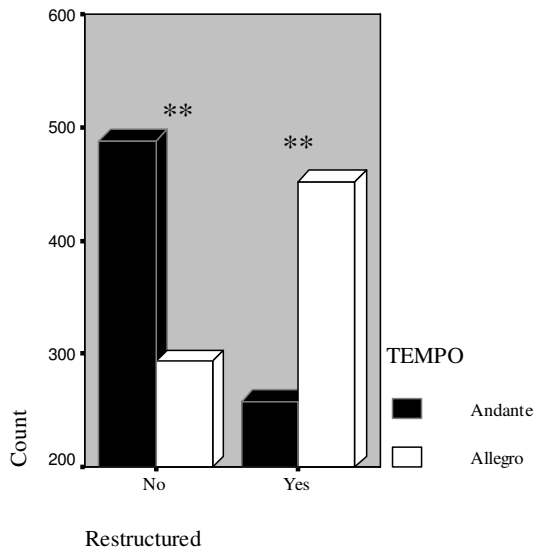
If the majority of these judgments indicated a restructuring as regards the correspondence pattern, we assigned this word 1 (yes), otherwise 0 (no). These judgments were analyzed with the help of a Pearson Chi-Square test, of which the results are shown in Figure 23. The graph clearly shows that the 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

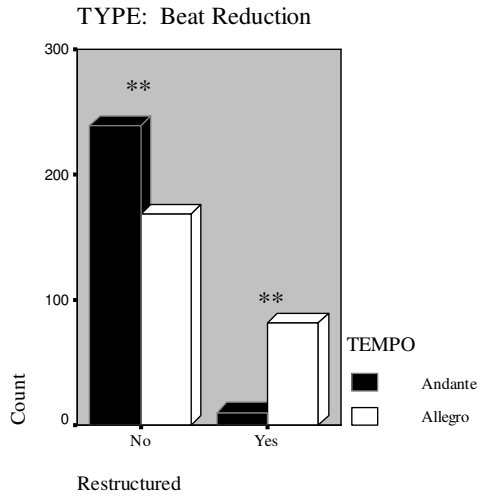
Count

		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



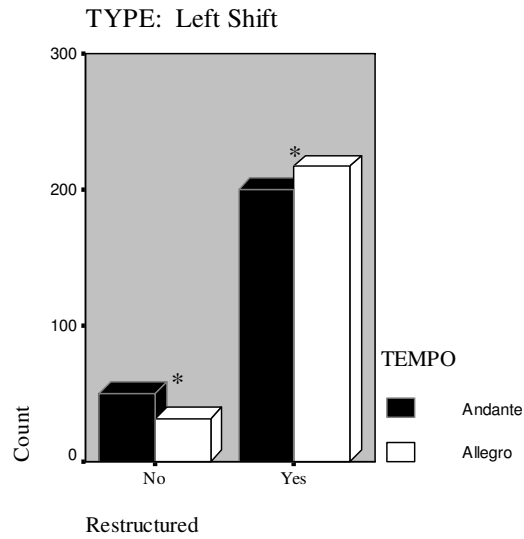
$$\chi^2 (1) = 67.781, p < 0.001$$

(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



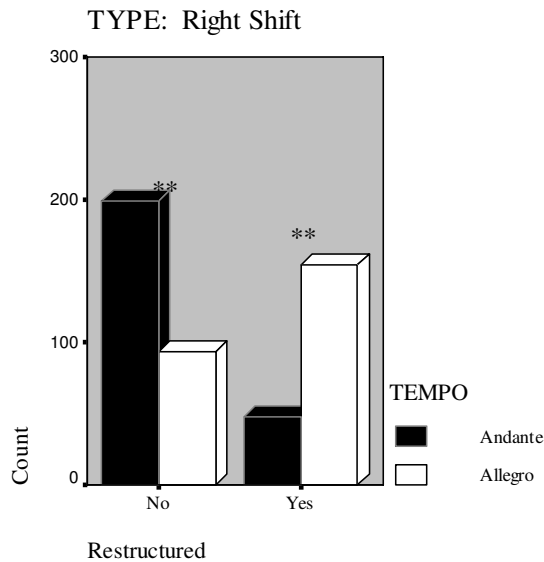
$$\chi^2 (1) = 4.642, p < 0.05$$

(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

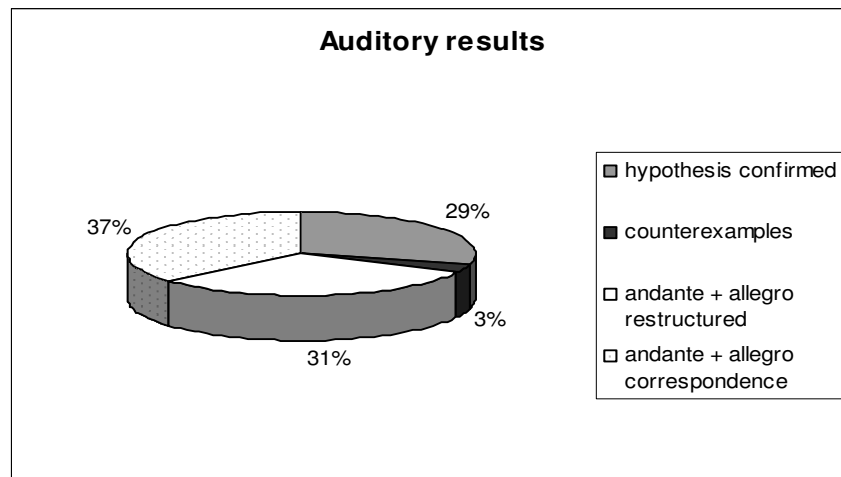
		TEMPO		Total
		1	2	
Count	0	199	93	292
	1	48	154	202

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.

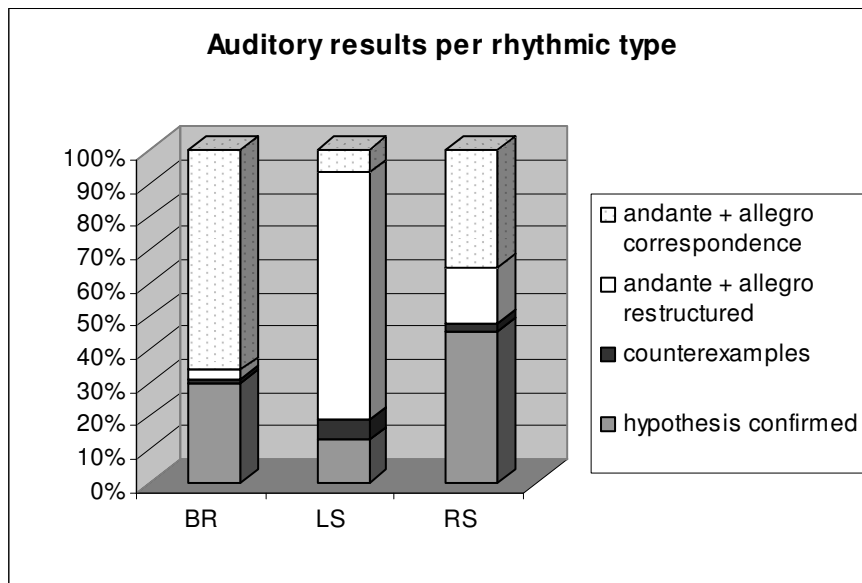
Figure 25a Pairs of the same word in andante and allegro



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

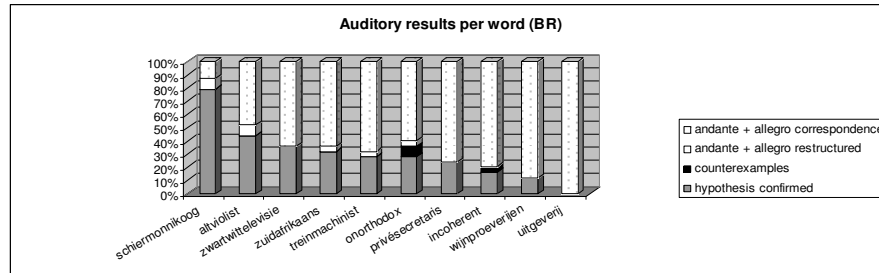


3.5.2.2. *Between-item variation*

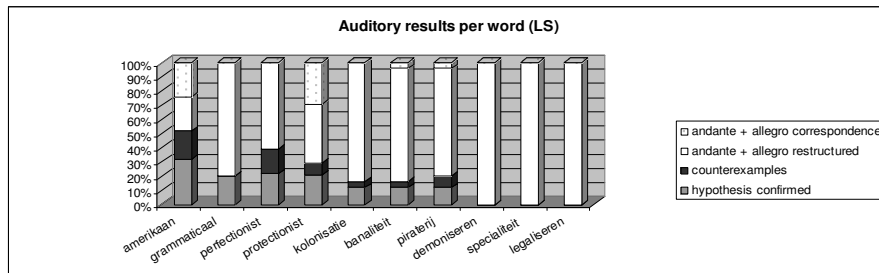
If we separate the results further and look at the behavior of the individual words in Figure 26a,b,c, we see that in the Left Shift words (Figure 26b) many individual words are responsible for the high number of counterexamples. Some of the Left Shift words, *demoniseren* ‘demonize’, *specialiteit* ‘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’.

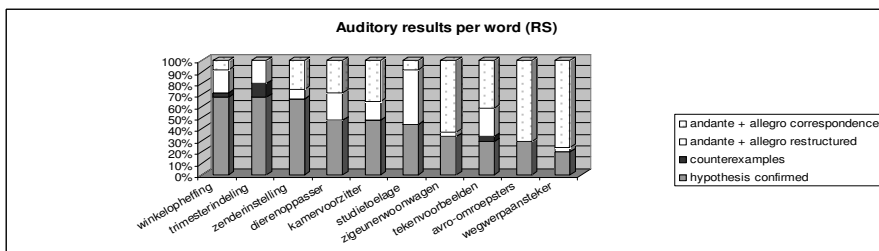
Figure 26 Pairs of the same word in andante and allegro, per word



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

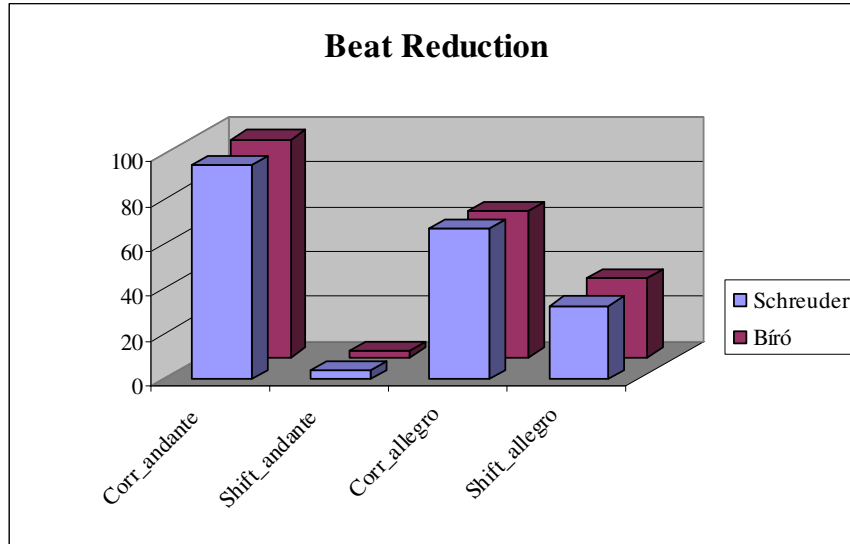
FTT=TR >> AFFIX-STRESS >> O-O CORR >> *ΣΣ >> PARSE-σ

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.¹⁵ Each rhythmic type thus has different local and global optima, and slopes, and will therefore show different frequency distributions. The

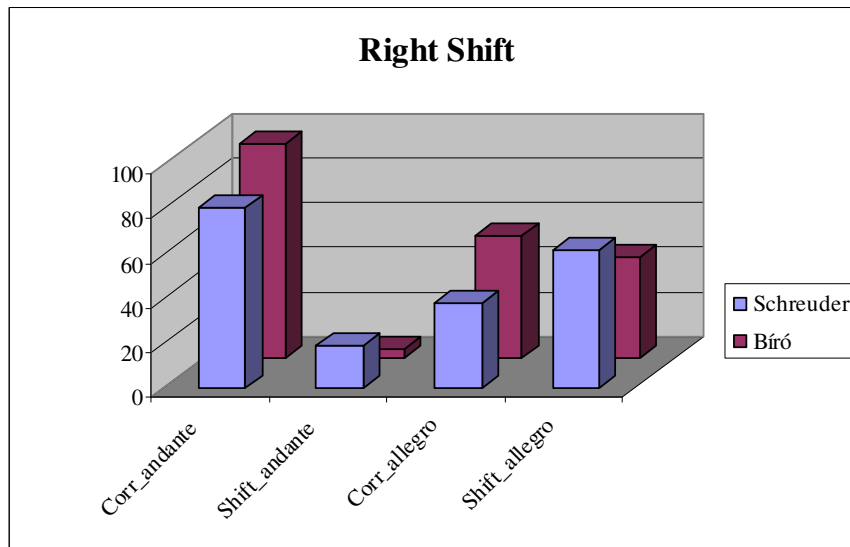
¹⁵ Bíró slightly redefines O-O CORR, which enables the simulation to count the number of violation marks assigned by this constraint to any candidate.

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

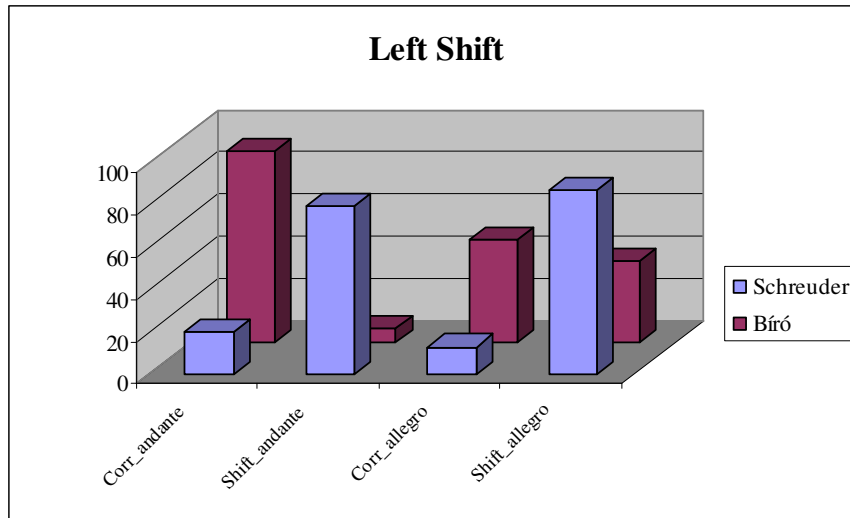
a.



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 *perfectionist* in andante tempo and as *perfectionist* 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

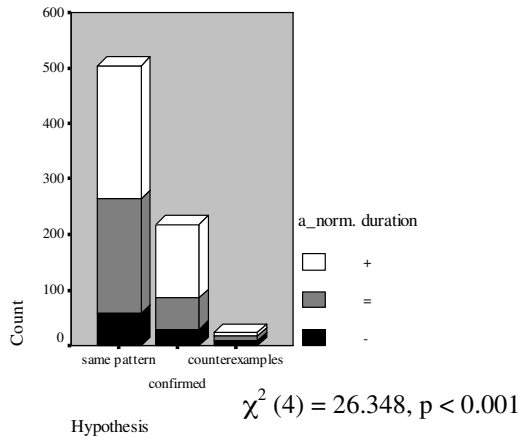
The only possible explanation for the outcomes in Table 23a,b is that the intrinsic values of the segments in the syllables might play a role here. This is because we took all data, andante and allegro, together in this test. Therefore, we separated the measurements of andante and allegro, and we subtracted 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

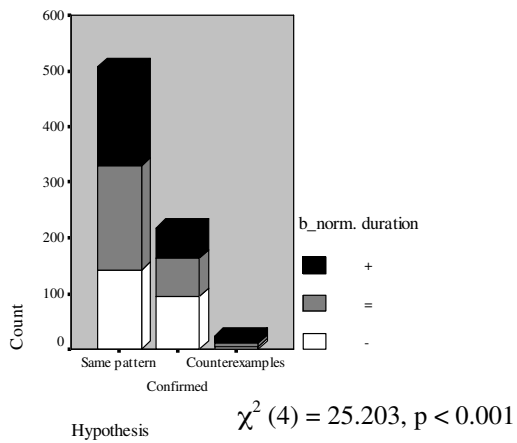
acoustic correlates duration, pitch and intensity are not the relevant correlates of secondary stress.

Figure 29

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



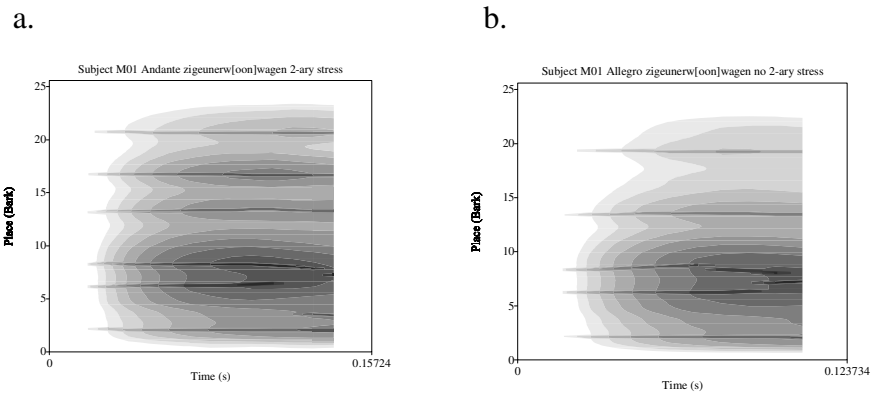
- b. Chi-Square: Andante values – Allegro values squared with auditory judgements (syllables b)



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.

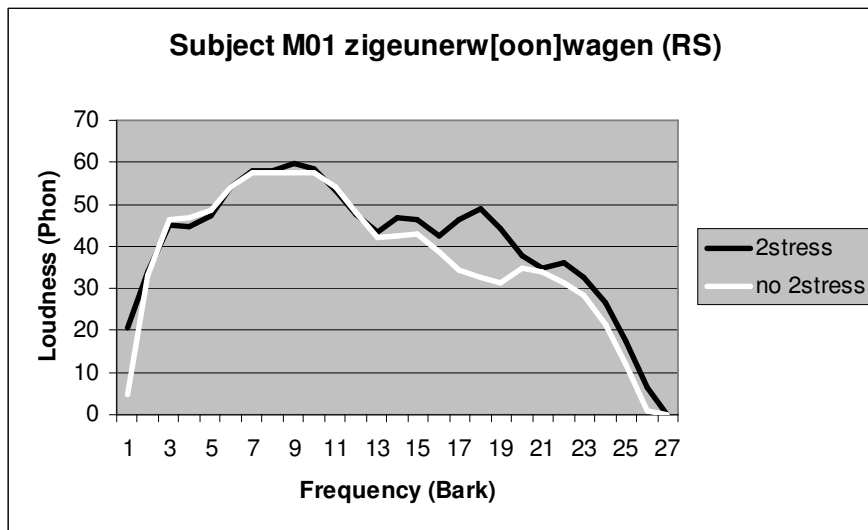
Figure 30 Cochleagrams of [o] in zigeunerw[oon]wagen 'gipsy trailer' (RS)



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.¹⁶ 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



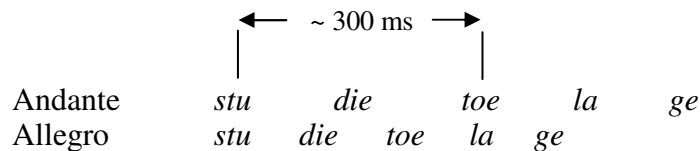
The white line in Figure 31 indicates the pattern of the allegro unstressed [o] in *zigeunerwoonwagen* and the black line indicates the pattern of the andante stressed [o]. We see increased loudness in the region of 13 to 20 Bark, which correlates with the most sensitive region of our ear.

¹⁶ 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^oelàge*, than in andante speech, e.g. only one syllable *die* in *stúdi^oelàge*. Figure 32 depicts this situation schematically.

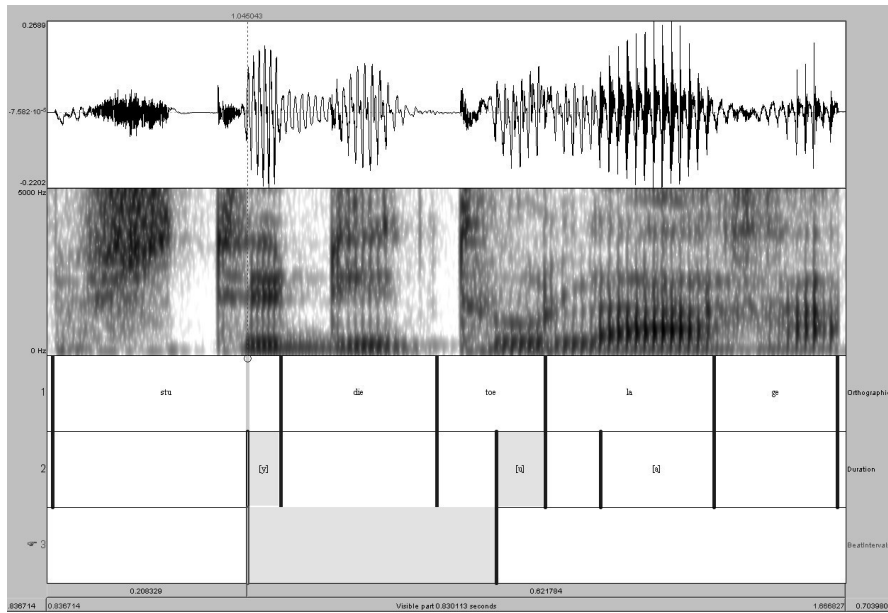
Figure 32 Beat Intervals



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

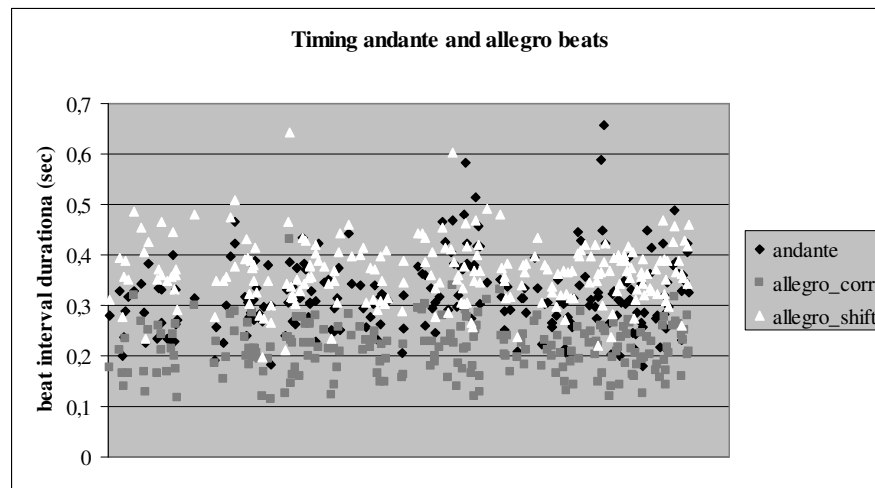


If indeed the perception of secondary stress shifts depends on rhythmic timing, the beat intervals between prominent syllables in *andante* and *allegro* speech are approximately equal. We measured the beat intervals between all possible stress placement sites for all word pairs which were perceived 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.¹⁷ The scatterplot in Figure 34 shows that

¹⁷ 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

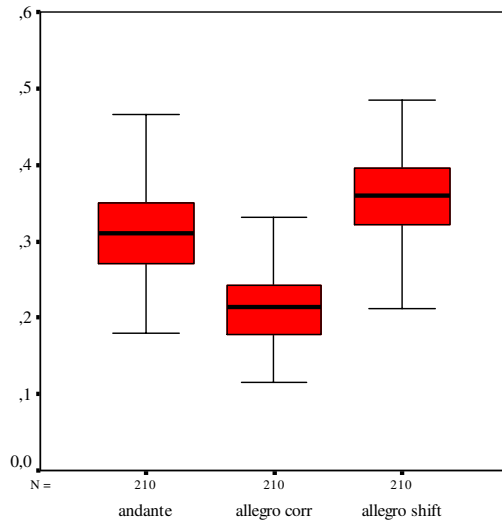
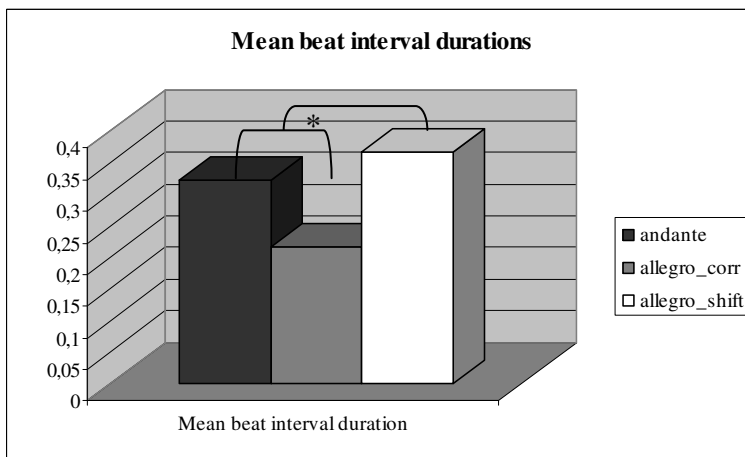


Figure 36 Mean beat interval durations of the three intervals



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.

