Institute of Phonetic Sciences,
University of Amsterdam,
Proceedings 22 (1998), 97-113

PERCEIVED PROMINENCE AND THE METRICAL-PROSODIC STRUCTURE OF DUTCH SENTENCES

Karijn Helsloot[*] *and Barbertje M. Streefkerk*

Abstract

In this paper we present a metrical-prosodic analysis of Dutch read aloud sentences, based on perceived prominences. Metrical-prosodic constraints are formulated which can be used as input for Text-to-Speech systems. A four level metrical grid representation is introduced, corresponding to four degrees of prominence, including no prominence. A distinction is made between input and output constraints. The former refer to the prosodic representation of lexically distinguished categories; the latter to sentence-level prosodic well-formedness. Constraints may be in conflict with each other. The conflicting constraints are relatively ranked in a constraint hierarchy. The higher ranked constraint will win at the cost of the lower ranked one.

1. Introduction

In the past ten to fifteen years different prosodic parsing models have been advanced for the purpose of text-to-speech (TTS) systems. But a close to natural speech sound realization of written texts continues to be an extremely difficult matter. With this paper, we do not intend to resolve the existing problems, but to provide new information on the subject which may lead to improvements of TTS systems.

In comparison with previously proposed prosodic parsing models for Dutch TTS systems (cf. Baart 1987, Dirksen & Quené 1993, and Quené & Kager 1993), we claim that our proposal is easier to implement, and at the same time more in accordance with the metrical variation that is observed in natural sentences. Instead of a set of metrical principles translating syntactic phrase structure rules, our model indirectly comprises syntactic relations by assigning different metrical values to verbs, nouns and modifiers. This metrical variation, in turn, in combination with a set of prosodic well-formedness constraints, gives rise to four degrees of prominence, each of which has to be translated in its proper acoustic values. The earlier-mentioned systems recognize only two degrees of prominence, no prominence, and accent prominence.

Since our corpus of analysis consists of newspaper sentences presented out of their context, the semantic-pragmatic distinction between given and new information is not included in our metrical parsing constraints.

[*] Studio Taalwetenschap Helsloot Verrips, Amsterdam.

It is generally assumed that prosodic parsing includes accentuation as well as phrasing. The perception task on the basis of which we define the metrical constraints, is restricted to information about accentuation (prominence), however. No constraints are thus formulated that refer to (different strengths of) boundaries.

2. The Phonetic Material

The speech material is selected from the Dutch Polyphone corpus. This corpus consists of 12500 newspaper sentences. A total of 5000 speakers were asked to read five sentences, to be recorded over the telephone (for more details see Damhuis et al. 1994). From this corpus, we took a random set of 50 sentences for the purpose of a metrical-prosodic analysis. Although the grammatical structure of the sentences varies, they are all declarative.

2.1 The Corpus of Dutch Sentences

The 50 newspaper sentences consist, on average, of 10.38 words per sentence, and the average number of syllables per sentence is 18.48, as shown in table 1. About half of the words are function words and the other half are content words. As expected, function words are perceived as being less prominent than content words.

Table 1: Number and means of words and syllables over 50 sentences and per sentence.

	total number	mean per sentence
words	519	10.38
content words	278	5.56
function words	233	4.66
rest words	8	-
syllables	924	18.48

2.2 Listening Experiment

The 50 sentences are part of a much larger set of sentences selected by Streefkerk for the purpose of a study on prominence perception (Streefkerk 1997). A first perception experiment, executed by Streefkerk, involved 500 sentences spoken by 50 male and 50 female speakers. Ten listeners, all students from the Humanities Faculty at the University of Amsterdam, were asked to indicate which words were realized with emphasis. The 500 sentences were presented in 4 random order sessions, which differed per listener, to compensate for possible learning effects. The first two sessions contained 150, and the last two sessions contained 125 sentences. The perception experiment was performed on a UNIX workstation, and the results of each listener were automatically stored. While hearing the sentence through closed headphones, the listeners saw the sentence on a monitor. Under each word, on the monitor, a button was placed. The subjects had to click on the button when a given word was perceived as being spoken with emphasis. To test the consistency of the listeners, 50 sentences were presented twice to each listener. This set of 50 sentences is used for the metrical-prosodic analysis.

An example of the perception results for one sentence is given in table 2. The sentence *De vliegtuigkaping werd tijdens de vlucht opgelost.* 'The airplane hijacking

was solved during the flight' was scored twice by the 10 listeners. For each word, the 20 judgements are added together, giving rise to a score between 0 (no mark) and 20 (all listeners marked this word twice as emphasized). We assume that the resulting scale of judgments is an indication of the involved degree of prominence: the higher the score the more prominent a given word is.

Table 2: Example of the results of the listening experiment. The table shows the cumulative judgments of the listeners and the resulting degrees of prominence.

Listener #	De	vliegtuigkaping	werd	tijdens	de	vlucht	opgelost.
1	0	1	0	1	0	1	0
.	.	.	.	.	.	.	.
9	0	1	0	1	0	0	1
10	0	1	0	1	0	1	0
Sum first	0	8	0	8	0	4	1
Sum second	0	8	0	8	0	6	3
Sum total	**0**	**16**	**0**	**16**	**0**	**10**	**4**

It should be mentioned that the listeners differ quite remarkably with respect to the number of emphasized words they perceive in one and the same sentence. While some listeners assign a mean number of 1 prominence per sentence (see e.g. listener 7, table 3), others assign 4 prominences (see e.g. listener 9, table 3). It strikes, for instance, that only 4 times all listeners (20 marks) agree that a certain word is emphasized. These facts argue in favor of a relative, instead of an absolute, metrical representation. That is, a prosodic analysis which rigidly translates the syntactic surface structure into prosodic constituents, as proposed for instance by Nespor & Vogel (1986), leads to an abundance of prosodic heads and boundaries which have no acoustic and perceptional correspondents.

Table 3 also shows the existence of a learning effect. The listeners 4, 6, 8 and 9 mark substantially more words as prominent during the second parsing than during the first one. In Streefkerk & Pols (1998) it is shown that the set of marked words in the first parsing is mostly a subset of the marked words in the second parsing. Although the differences within and between listeners are rather strong, we still consider the cumulative judgements to be a useful alternative for prominence labeling of the speech material.

Table 3: Number of prominence judgments per listener after first and second parsing.

Listener #	1	2	3	4	5	6	7	8	9	10	sum
First 50	71	50	160	165	135	132	50	109	156	172	1200
Second 50	71	51	165	202	130	211	50	149	209	158	1396

In table 4 we present the distribution of prominence marks according to a threefold distinction: (i) 0 marks, (ii) from 1 to 10 marks, and (iii) from 11 to 20 marks. The mean number of 0 marked words is 4.88 per sentence. This equals the amount of function words per sentence, as illustrated in table 1. The sum of marked words gives rise to a mean of 5.5 words per sentence. This equals the amount of content words per sentence (see table 1).

Table 4: Distribution of prominence marks between 0 and 20.

Prominence	Total	Mean per sentence
11-20	136	2.72
1-10	139	2.78
0	244	4.88
Total	519	10.38

A final observation must be made here with regard to the listening experiment. Since the listeners were asked to indicate emphasized words, and not emphasized syllables, and they were *not* asked to indicate degrees of emphasis, the results do not give us information about (i) the location of the emphasis within the word, and (ii) the presence of weakly stressed words/syllables. With respect to (i), in general, the lexically stressed syllable of the word is also the syllable actually realized with prominence. Only in a very few cases, lexical stress shifts can be observed. With respect to (ii), unfortunately the prominence values of polysyllabic function words and of secondary stresses within relatively long content words are not perceptually tested. But in a similar experiment on a different set of sentences from the Polyphone corpus, the task included perceived prominence of words versus perceived prominence of syllables. The listeners indicated a mean number of 2.9 words, but a mean number of 5.1 syllables per sentence as being prominent (Streefkerk et al. 1997). This result seems to confirm that weakly stressed syllables are indeed perceived when the perception task is formulated differently.

Other perception tests in which degrees of prominence were asked for, also indicate that listeners are able to differentiate between unstressed, weakly stressed and strongly stressed syllables (see Helsloot 1993, 1995). In addition, the acoustic signals as well as careful listening of short trunks of the Dutch sentences reveal the presence of different stress degrees.

3. The Metrical-Prosodic Analysis

With the results of the listening experiment on the one hand, and rather elaborated theories of prosodic phonology on the other (cf. among others, Selkirk 1980, 1986), we hypothesized that a mapping of the two would be feasible. The six-level organization assumed by prosodic phonology, i.e. the syllable, the foot, the prosodic word, the phonological phrase, the intonation phrase and the phonological utterance, appeared to be a far too rich as well as a far too rigid system, however. Instead of six levels, only four levels could be distinguished. And the classical assignment of prosodic constituent structure to sentences based on morphological and syntactic information gave rise to absolutely determined heads and constituents which very often were not encountered in the perception results. A relational-based metrical grid representation, as initially proposed by Liberman & Prince (1977), extended by restrictions on the number of hierarchical levels, instead, allows for a more adequate analysis of the metrical structure of the material.

3.1 Prosodic Input Constraints

Since we are dealing with read sentences and not with spontaneous speech, almost all syllable inputs are properly realized in the output. Syllable deletion or syllable insertion, well-known phenomena in Dutch spontaneous speech (see Kuijpers & van

Donselaar 1998), occur in just a few cases. In the 50 sentences, we found four instances of syllable deletion, and two instances of syllable insertion. The involved vowel is always a schwa:

(1)

syllable deletion		*syllable insertion*	
gist[ə]ren	> gistren	merkt >	mer[ə]kt 24
'yesterday'		'(he) notices'	
vriend[ə]lijke	> vrienlijke	half uur >	hall[ə]f 38
'friendly'		'half an hour'	
verzek[ə]ringsagent	> verzekringsagent		
'insurance agent'			
Ned[ə]rland	> neland		
'the Netherlands'			

Generally put, in a sequence of two or more unstressed syllables, the left-most schwa tends to be deleted, and in pre-boundary position, or in a sequence of adjacent stresses, a schwa is inserted if permitted by the segmental environment. In other words, rhythmic lapses and clashes are possibly resolved at the syllable level. Obviously, a TTS system must include these rhythmically-driven syllable deletions and insertions.

Apart from these insertions/deletions, input syllables are realized at the surface. This observation is metrically represented by a mark on the lowest metrical grid level:

(2) *Syllable Constraint*
All syllables receive a level-1 mark on the metrical grid.

Thus, the sentence in (3) is initially parsed into a sequence of level-1 marks:

(3)
x x x x x x x x x x
De ontspoorde trein ramde het taluud.
'The derailed train rammed the bank'

3.2 Function Words

In our corpus a total of 233 function words occur, comprising Determiners, Auxiliaries/Modals/Copulas, Prepositions, Possessive Pronouns, Complementizers, Personal Pronouns, Reflexive Pronouns, (Anaphoric) Demonstrative Pronouns, and Conjunctions. Of these function words, 216 are monosyllabic, and 17 polysyllabic. Table 5 presents the relevant distributions.

Table 5: Number and means of monosyllabic and polysyllabic function words regarding prominence degrees.

Function words	*total*	*0 Prom*	*%*	*1< Prom ≤10*	*%*	*10 < Prom ≤20*	*%*
Monosyllabic	216	205	95	11	5	-	-
Polysyllabic	17	12	70	2	12	3	18

The listeners perceived 205 monosyllabic function words as bearing no prominence at all. Eleven monosyllables were perceived as bearing a (very) low degree of prominence (1< Prom ≤ 10). Three distinct explanations for this low degree of prominence can be given: (i) the monosyllabic function word occurs in absolute

sentence-initial position, (ii) the monosyllable is prominent in order to avoid a rhythmic lapse, and (iii) the monosyllable receives prominence because it is part of a slowly read speech string in which all syllables are realized with prominence. Some examples (the numbers following the sentence fragments correspond to the number of marks assigned by the listeners to the separate words):

(4)	i.	**Ik** was hier al	**1**.6.2.3
		'I was here already'	
		In feite is Nederland	**1**.19.0.2
		'In fact the Netherlands are'	
	ii.	...bevatten **dat** ze had gewonnen	0.**1**.0.0.4
		'...grasp that she had won'	
		...wachten **op** het moment...	1.**1**.0.10
		'...wait for the moment'	
	iii.	...van algemene rouw **ten** gevolg	0.11.14.**1**.2
		'...of general mourning as a result'	

The explanations (i) and (iii) are pragmatical in nature, and possibly also speaker-dependent. It is certainly not the case that sentence-initial monosyllabic function words tend to be realized with prominence, or are systematically perceived as being stressed. The decrease of speaking rate, observed in a few readings, is neither grammatically determined. That is, slow reading must not be incorporated in the basic prosodic parsing of a TTS model. By contrast, explanation (ii) is, like the above-mentioned phenomena of syllable deletion/insertion, an example of rhythmic readjustment: a stress is added in order to avoid a rhythmic lapse. In section (4.2.2), this grammatically determined rhythmic readjustment is formalized in terms of a prosodic output constraint.

As is indicated in the table, 95% of the monosyllabic function words are perceived as completely stressless. This amounts to the following TTS prosodic input constraint:

(5) *Function Word Constraint I:*
Monosyllabic function words do not receive a grid mark on level-2 or higher.

Regarding polysyllabic function words, the following constraint is proposed.

(6) *Function Word Constraint II:*
The head syllable of a polysyllabic function word receives a level-2 mark in the metrical grid.

Although the table indicates that 70% of all polysyllabic function words in the corpus were perceived as bearing no prominence at all, we nonetheless retain that polysyllabic function words are more prominent than monosyllabic ones. First of all, in citation form, a speaker of Dutch will indicate one of the syllables of a functional polysyllable as bearing word stress. Secondly, as said in the introduction, because of the formulation of the perception task listeners focussed on words realized with emphasis. Clearly, function words generally are not realized with emphasis, although they may be realized with a low prominence degree. And thirdly, the acoustic representations of the read sentences clearly indicate that weakly stressed syllables have particular acoustic properties. Although an acoustic analysis is outside the scope of this paper, a TTS model should translate weakly stressed syllables in order to get close-to-natural realizations. In the models for Dutch, known to us, weak or secondary stress is completely neglected.

The polysyllabic function words perceived as being realized with prominence are mostly emphasized prepositions. In other words, the prepositions received contrastive focus:

(7) a. De vliegtuigkaping werd *tijdens* de vlucht opgelost.
'The airplane hijacking was solved *during* the flight' (and not *after* the flight)

b. Mijn verzekeringsagent woont *tussen* de medisch specialisten in Beugen.
'My insurance agent lives *amid* the medical specialists in Beugen' (and not in another neighborhood, as you would expect)

Obviously, this contrastive focus must be accounted for by another constraint than *Function Word Constraint II*.

3.3 Content Words

Content words are stressed, but function words not. This prosodic distinction is uncontroversial in the phonological literature, as well as in the TTS models that have been proposed for Dutch, and for other stress-based languages (see e.g. O'Shaughnessy 1976, Baart 1987). Indeed, in our corpus of read sentences, listeners perceived prominence on many if not all nouns, verbs, adjectives and adverbs. In table 6, the exact results are given. Adverbs and adjectives are always perceived as being realized with prominence. Nouns, in general, too, with the exception of three instances (2% *vs.* 98%). And verbs are mostly perceived as being realized with prominence, although the rates are more balanced (38% no prominence vs. 62% prominence).

Table 6: Number and means of content word categories regarding prominence degrees.

Content words	*Total*	*0 Prom*	*%*	*1 < Prom ≤ 10*	*%*	*10 < Prom ≤ 20*	*%*
Nouns	143	3	2	71	50	69	48
Verbs	50	19	38	25	50	6	12
Adverbs	33	-		17	49,5	16	48,5
Adjectives	46	-		18	40	28	60

This prosodic property of content words leads to the following constraint:

(8) *Content Word Constraint I:*
The head syllable of a content word (either monosyllabic or polysyllabic) receives a level-2 mark on the metrical grid.

As shown by the figures in table 6, verbs are prosodically less prominent than the other word categories. This tendency characterizes west-Germanic languages like English and Dutch. As reported by Baart (1987:57), scales of accentability for content word classes in English place main verbs lower than nouns, adjectives and adverbs (cf. Lea 1979, O'Shaughnessy & Allen 1983). For Dutch, Kruyt (1985) argues that verbs have a lower accentability degree than nouns and adjectives, but a slightly higher degree than adverbs. In our corpus, the position on the scale of accentability for Dutch adverbs, as proposed by Kruyt, cannot be confirmed. Adverbs are mostly perceived as highly prominenced. The following constraint formalizes our findings, as reported in table 6:

(9) *Content Word Constraint II:*
The head syllable of a noun, adjective or adverb receives a level-3 mark on the metrical grid.

The constraints proposed until now assign the following metrical grid representation to the sentence in (10).

(18)

```
          x       x                 x
          x       x     x           x
  x   x   x   x   x     x   x   x   x x
De ontspoorde trein ramde het taluud
'The derailed train rammed the bank'
```

Before addressing instances deviating from *Content Word Constraints I* and *II*, we first address the metrical input representation for compounds and complex verbs.

3.3.1 Compounds

Dutch is a language with a highly productive system of compounding. Almost a fifth of all content words in our corpus consist of compounds. Although the number of nominal compounds exceeds by far the number of adjectival, verbal and adverbial ones, all four categories allow for productive compounding. The lexical or morpho-phonological properties of Dutch compounds have given rise to a variety of metrically rich and complex representations (see Visch 1989, Booij 1995). We suggest that the metrical representation of compounds at the phrase-level, however, can be reduced to two. One input representation for compounds with adjacently stressed syllables, and one for compounds with non-adjacently stressed syllables. All disyllabic compounds belong to the former type. Metrically larger compounds belong to either the former or the latter type. The relevant prosodic input constraints are graphically formalized and illustrated in (11).

(11) <u>*Compound Constraints*</u>*:*

Class I Compounds

```
x          x
x  x   >   x
x  x  (x)  x  x  (x)
```

wíelrènner (N) 'cyclist'
kléurrìjk (A) 'colorful'
ópzìen (V) 'look up'

Class II Compounds

```
x             x
x     x   >   x     x
x  x  x  (x)  x  x  x  (x)
```

ménsenschùw (N) 'shyness'
èensgezínd (A) 'unanimous'
úitgerùst (V) 'in condition'

Class I compounds are thus treated similarly to simplex content words. A TTS model just needs to know which of the two syllables is the head of the word. Whether the word forms a compound or not is irrelevant for the purpose of an adequate metrical TTS representation. In (12) a sentence is given with a Class I and a Class II compound.

(12)

```
      x           x           x           x
      x           x  x        x     x     x
  x   x   x  x    x  x  xx  x  x  x  xx  x  x  x   x  x
De <u>werknemers</u> van <u>Unilever</u> in India hebben het moeilijk
'The employees of Unilever in India are in trouble.'
```

3.3.2 Complex Verbs

Dutch is rich of complex verbs, formed by a verbal stem preceded by a prepositional particle. When inflected, the verb and particle are separated from each other, such that the particle occurs in clause-final position and the verbal stem in the second position of the main clause. Particle and stem form a compound in participles and in infinitives, except for infinitive constructions in which the infinitive marker *te* occurs. In the latter case, the word order is particle + infinitive marker *te* + verbal stem. In other words, there are two constructions leading to *surface compounds* and two constructions leading to *surface separation*.

Lexically, most verbal compounds have main stress on the particle: lángskomen 'come by', óphalen 'fetch'. With respect to prominence perception, the following tendencies are observed in our corpus: (i) the *surface compounds* are always perceived as prominent (e.g. 13a), (ii) the verbal stems in the separated forms are never perceived as being prominent (e.g. 13b), (iii) the particles in the separated forms are always perceived as being prominent in the particle+*te*+verb infinitive constructions (e.g. 13c), but (iv) the clause-final particles in inflected forms are only perceived as prominent if they do *not* occur in sentence-final position, i.e. when they occur in the non-final clause of the sentence (e.g. 13d); in sentence-final position they are not perceived as prominent (e.g. 13b). (Verbal stem and particle are marked by italics, and the prominent word by boldface).

(13) a. Wij hebben de beukeboom die in de tuin stond, ***omgehakt***.
'We have cut down the beech tree that we had in the garden.'

b. Ajax *ging* in de kwartfinale van het eindtoernooi kansloos *onderuit*.
'Ajax didn't have a chance in hell in the quarter-finals of the final tournament.'

c. De voetballer is verplicht om zijn contractuele verplichtingen ***na te komen***.
'The soccer player is obliged to observe his commitments by contract.'

d. Na morgen *zet* het vriendelijke weer zich ***voort*** en zijn de buien verdwenen.
'After tomorrow the friendly weather will continue and the showers will disappear.'

On the basis of these heterogeneous findings it is not immediately clear how to represent complex verbs metrically. Two options are available: (i) surface compounds as well as separated complex verbs are subject to the *Compound Constraints*, or (ii), *surface compounds* are subject to the *Compound Constraints*, but separated forms are subject to the relevant *Function Word Constraints* and *Content Word Constraints*. For a TTS system the second option is easier: no syntactic analysis is required, the verbal stem is identified as a simplex content word, and the particle as a mono- or polysyllabic preposition. However, in order to avoid that the particle in the particle+*te*+verb construction will be identified as a stressless element, a level-2 mark must be assigned to these particles in the metrical grid. Furthermore, the separated particle in clause-final (but not sentence-final) position must receive its proper metrical interpretation. Although we applied the second option in our analysis, we do not have a solution for the metrical behavior of the sentence-medial separated particle.

3.4 Lexical Modifiers

A final metrical input constraint will be presented now. The results of the perception task show that in most cases in which an argument is modified by a word (and not by a phrase), this modifier is perceived as bearing more prominence than the modified

word. Relevant sequences are adjective-noun, verb-adverb, adverb-verb, and adverb-adjective sequences. A couple of examples are given below.

(14)	kwaliteitsvolle marathonloper 'a quality marathonracer'	11.6	Adj-N
	vanuit rijdende auto's 'from riding cars'	0.12.1	Adj-N
	riskeerde bewust 'risked consciously'	0.19	V-Adv
	vaak meet men 'one often calculates'	10.0.0	Adv-V-Pro

Of the 35 adjective-noun sequences occurring in the corpus, 30 are perceived as having a higher degree of prominence on the adjective than on the noun. This amounts to 86% of all adjective-noun sequences. In the case of adverb-verb or verb-adverb sequences there are no instances at all in which the verb is perceived as bearing a higher degree of prominence than the adverb. On the basis of these facts, we formulate the *Modifier Constraint*, which assigns a level-4 mark to a lexical modifier.

(15) *Modifier Constraint*: Each lexical modifier receives a level-4 mark.

Regarding the 5 instances in which the adjective has no higher degree of prominence than the noun, it is observed that in 4 instances the prominence degrees of the two words are either identical or almost identical:

(16)	a.	potentiële betrokkenen 'potentially involved (persons)'	9.9
	b.	verschillende auto's 'various cars'	8.9
	c.	om half zes 'at half past five'	0.6.8
	d.	algemene rouw 'general mourning'	11.14

The only clear-cut iambic, or weak-strong pattern is realized on:

(17)	van de oude fietsroutes 'of old cycle routes'	0.0.3.12

Broad focus, lexicalization, high frequency words, non-finality, and maybe other explanations might be advanced as underlying this marked pattern, but all these properties equally characterize one or more of the other adjective-noun sequences giving rise to a trochaic pattern. The tendency to emphasize the adjective and not the noun prevails very clearly. In fact, the phrases in (16) and (17) also allow for a trochaic, or strong-weak realization.

Negative particles and deictically used demonstrative pronouns are also realized with a high degree of prominence. These elements too are subject to the *Modifier Constraint.*

The entire set of metrical input constraints gives rise to the following representation for one of the sentences from the corpus.

(18)

```
          x
          x       x                   x
          x       x     x             x
  x   x   x   x   x     x   x   x   x   x
De ontspoorde trein ramde het taluud
```

'The derailed train rammed the bank.'

4. Prosodic Output Constraints

The metrical input constraints do not cover correctly all the metrical patterns that are observed throughout the sentences. Two metrical patterns in particular require a specific account, referring to the beginnings of sentences and the endings of sentences. It strikes that sentence-initial content words (except verbs) have a higher degree of prominence than following content words. In the case of sentence-initial adjective-noun or adverb-verb sequences no intervention is needed, since lexical modifiers receive a higher grid mark by means of the *Modifier Constraint* presented in 3.4. But when the first constituent contains a non-modified noun and the immediately following constituent is either a modifier or a non-modified noun, ill-formed metrical patterns arise. That is, the sentence-initial noun has either a lower grid-level or the same grid-level as the following content word:

(19)

```
        x                       x   x
    x   x                       x   x
    x   x                       x   x
  ([N] [AN]                   ([N] [N]
```

The examples in (20) illustrate the actually observed prominence patterns. The prominence degrees refer to the words in italics:

(20)

a.	De *bewijslast* tegen de *verdachte* 'The onus of proof on the suspect'	19.16
b.	De *verzekeraars* hebben het *moeilijk* 'The insurers have difficulties'	17.17
c.	Zijn *reputatie* als *kwaliteitsvolle marathonloper* 'His reputation as a quality marathonracer'	16.11.6

To ascertain that in sentence-initial position a higher degree of prominence is realized than on the following content word, the *Sentence-Initial Constraint* is assumed to operate:

(21) *Sentence-Initial Constraint:*
The first level-3 mark in the sentence receives a level-4 mark.

The explicit reference to words with a level-3 grid mark prevents verbs to be subject to this constraint. In fact, a sentence like *Hij eet een appel* 'He eats an apple' has the highest prominence degree on the noun *appel* and not on the verb *eet*.

First name plus surname, title plus surname, and dates form exceptions to this pattern: the highest prominence is on the second content word:

(22) Pater Groenewegen 4.15
'Father Groenewegen'

in de maand februari 0.0.0.17
'in the month of February'

Sentence-finally, a similar trochaic pattern is mostly observed. That is, if the input gives rise to a strong-strong sequence (level-3 plus level-3, or level-3 plus level-4), a strong-weak output is realized. Since sentence-final words are typically perceived as weakly prominent, the relevant constraint is defined as follows:

(23) *Sentence-Final Constraint*:
The final level-3(4) mark in the sentence is deleted.

The fact that the base position for verbs in Dutch is sentence-final, causes that many sentences in our corpus give rise to a sentence-final trochaic pattern. Verbs have a low grid level, as illustrated by the sentence-final strings in (24) (the prominence marks refer to the words in italics):

(24)	a.	aan het *orgel gewijd.*	14.2
		'dedicated [his life] to the organ	
	b.	een *shuttle gelanceerd.*	15.0
		'a shuttle launched'	
	c.	en zijn de *buien verdwenen.*	10.1
		'and the showers will disappear'	

The constraint is required, however, in order to account for outputs like those in (25):

(25)	a.	bij een *winkelcentrum* te *Houten.*	11.3
		'at a shopping centre in Houten'	
	b.	op het *moment* van *gloreren.*	10.5
		'at the moment of glory'	
	c.	vanuit *rijdende auto's*	12.1
		'from riding cars'	

The Sentence-Final Constraint reduces the final strong prominence to a weak prominence by grid-mark deletion, indicated by the angled brackets:

(26)
```
x    <x>
x     x
x     x
```

An example is given below.

(27)
```
            x          <x>
            x           x
x  x   x  x   x   x  x  x
op het moment van gloreren
```

4.1 Prosodic Maximality

Violations of the *Sentence-Final Constraint*, giving rise to an iambic pattern, do occur, however. Of the 50 metrically analyzed sentences, 7 have an iambic ending. Rhythmic alternation and domain maximality are the underlying reasons here. To start with the latter phenomenon, domain maximality, a sentence-final iamb is created when the preceding content word is 'too far away' from the final word. More precisely, the trochee must be realized within a maximum number of syllables. If the

syllables intervening the head syllables of the final and pre-final content word exceed the number of four, the input pattern is left unchanged. Two examples are given below (head syllables are given in boldface, and the prominence marks refer to the words in italics):

(28)	a.	nauwelijks *bevatten* dat ze had *gewonnen* 'hardly grasp that she had won'	0.4
	b.	bij de *bevolking* tegen de *politiek* 'by the population against politics'	9.14

Between the two head syllables *vat* and *won* in (28a) occur five syllables, and between the head syllables *vol* and *tiek* in (28b), six syllables. In Helsloot (1995), a chapter is dedicated to prosodic phrase maximality in Italian. Roughly said, in Italian the maximal phrase includes six to seven syllables, parsed into three feet. The final parts of the strings in (28) give rise to such maximal phrases: [dat ze had gewonnen] and [tegen de politiek]. In fact, in addition to the sentence-final verb *gewonnen* the complementizer *dat* is perceived by the listeners as having low prominence, and in (b), the disyllabic preposition *tegen*, although not perceived as bearing prominence, is expected to be realized by a low degree of prominence, on the basis of our *Function Word Constraint II*. The trochaic pattern required by the *Sentence-Final Constraint* cannot be realized because of maximality conditions on metrical phrasing. In the prosodic framework we are presenting here, phrases and boundaries are not explicit entities, however. The observed violation of the *Sentence-Final Constraint* can also be accounted for by a constraint on rhythmic alternation, which refers to alternating stress degrees instead of phrase size.

4.2 Rhythmic Alternation

Prosodic well-formedness mainly involves rhythmic alternation: at the phrasal level stresses must be distributed in accordance to principles of recurrence and hierarchical ordering (cf. amongst others, Prince 1983, Hayes 1984, Halle & Vergnaud 1987, Helsloot 1995). Recurrence is generally assumed to be of a binary nature: a strong beat is followed by a weak beat. This alternation applies at hierarchically ordered levels of organization. In other words, binary alternation is observed from the smallest to the largest domains of prosodic organization. Metrical grids insightfully represent this organization. Well-formed rhythmic alternation is shown in (29a), ill-formed alternations in (29b) and (29c).

(29)

a.	b.	c.
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 x x x x x

4.2.1 Clash Avoidance

Obviously, the words of a language when grouped into phrases do not always give rise to well-formed rhythmically alternating patterns. The input may show up with a so-called stress clash (as illustrated in 29b). Since stress clashes are hard to produce physically, they generally do not occur in the actual realization of the phrase. A clash can be resolved in various ways. By means of (i) deletion of one of the stresses, (ii)

movement of one of the stresses, or (iii) lengthening of the interval between the stresses. Whether one resolution or the other is selected depends on the linguistic context. In Dutch, all three resolutions are found. And it is mostly the stress on the right (i.e. the second stress) that will be modified. In case a clash exists between three stresses, things are becoming more complex.

In section 4.2, we referred to instances violating the *Sentence-Final Constraint*. Instead of a final trochee, an iamb is realized. The violation is due to stress-clash avoidance. Consider the metrical representation of the sentence in (30).

(30)

```
                                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
  De rechter spreekt van een onthutsend gebrek aan plichtsbesef.
  'The judge speaks of a bewildering lack of sense of duty.'
```

The *Sentence-Final Constraint* deletes the sentence-final level-3 mark. However, the perceived prominence pattern is 5 on *gebrek* and 11 on *plichtsbesef*. A clear iambic pattern. The representation in (30), in fact, gives rise to a clash with the level-4 mark of the preceding word, *onthutsend*.

In our corpus, about ten instances are found in which a sequence of three content words gives rise to a double stress clash. On the basis of the perceived prominences, we formulate a *Clash Resolution Constraint* which requires left-headedness at the highest level of organization.

(31) *Clash Resolution Constraint:*
In a sequence of three adjacent level-3/4 marks left-headed alternation is realized.

```
x     x     x           x
x     x     x     >     x           x
x     x     x           x     x     x
```

On the basis of this constraint, the above sentence is represented as follows:

(32)

```
        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
  De rechter spreekt van een onthutsend gebrek aan plichtsbesef.
  'The judge speaks of a bewildering lack of sense of duty.'
```

That is, the level-3 mark on *gebrek* will not be realized in the output.

The two relevant constraints are ranked with respect to one another: the *Clash Resolution Constraint* dominates the *Sentence-Final Constraint*, which explains the fact that the latter is violated.

Other examples from our corpus subject to the *Clash Resolution Constraint*, but not in sentence-final position, are given in (33).

(33) a.

```
              x     <x>
              x     <x>         x
  x           x     x           x
  x  x        x     x  x        x     x
  zitten vol groene vruchten                  0.5.3.8
  'are full of green fruits'
```

b.

```
        x        <x>      <x>
        x        <x>       x
        x         x        x          x
  x     x   x     x   x    x    x     x
  aanzienlijk minder mensenschuw    16.1.4
  'substantially less shy'
```

The *Modifier Constraint* assigns a level-4 mark to the modifiers *vol* and *groene* in (33a), and *aanzienlijk*, *minder* and *mensenschuw* in (33b). Since these level-4 marks are not separated from each other by intervening level-3 and level-2 marks, they are subject to the *Clash Resolution Constraint*. This latter constraint outranks the *Modifier Constraint*, i.e., it is satisfied at the expense of the *Modifier Constraint*.

4.2.2 Lapse Avoidance

The rhythmic counterpart of the clash is the likewise ill-formed *lapse*, illustrated in (29c) above. The metrical input may give rise to a relatively long sequence of syllables that are only characterized by level-1 marks. Such a pattern gives rise to a monotonous realization. Generally, such patterns are avoided in natural speech. The *Lapse Resolution Constraint* accounts for this avoidance.

(34) *Lapse Resolution Constraint*:
In a sequence of more than three level-1 marks a level-2 mark is assigned to the central syllable with a full vowel.

In section 3.2, we mentioned that monosyllabic function words are sometimes perceived by the listeners as being realized with a low degree of prominence. The avoidance of a rhythmic lapse explains this prominence (the mark between square brackets is added):

(35)

```
        x      [x]          x                    x
        x      [x]          x                    x
  x     x   x   x   x   x    x    x     x    x    x
  bevatten dat ze had gewonnen     wachten op het moment
  'grasp that she won'             'wait for the moment'
```

As said before, listeners were asked to indicate which words in the sentences were realized with emphasis. Word-internal weakly stressed syllables are thus not separately marked. Another kind of perception test is needed in order to get judgements about the presence or absence of weakly stressed syllables. However, TTS models which do not translate weak (or level-2) stresses in acoustic terms, will produce monotonous realizations of such weakly stressed intervals. The *Lapse Resolution Constraint* prevents such realizations to occur.

5. Conclusions

The perception-based analysis and the metrical-prosodic analysis of the Dutch read-aloud sentences give rise to a metrical grid representation containing four distinct levels of prominence. The lowest level corresponds to the syllable level; the second level to weakly stressed syllables which either belong to a function word, to a non-head syllable of a long content word, or to the head syllable of a verb; the third level

corresponds to the head syllable (lexically stressed syllable) of nouns; and the fourth level to the head syllable of adjectives and adverbs, and to negative particles and deictically used demonstrative pronouns. This representation is accounted for by a set of prosodic input constraints: the Syllable Constraint, the Function Word Constraints I and II, the Content Word Constraints I and II, the Compound Constraints, and the Modifier Constraints.

In addition, a number of sentence-level output constraints are formulated which account for the rhythmic well-formedness of the sentences: the Sentence-Initial Constraint, the Sentence-Final Constraint, the Clash Resolution Constraint, the Lapse Resolution Constraint.

The prosodic output constraints are often in conflict with the prosodic input constraints. The surface realizations indicate that the former are higher ranked in the constraint hierarchy than the latter. Output constraints themselves may also be in conflict with one another. For instance, the Sentence-Final Constraint can be violated by the higher ranked Clash Resolution Constraint.

The metrical grid representations resulting from the constraints match extremely well with the presence versus absence of prominences as perceived by the listeners: (i) of the 275 words perceived as prominent only eleven do not receive a proper metrical representation (see the prominence marks assigned to monosyllabic function words in table 5), (ii) and only one word with a level-3 mark (or higher) is not perceived as prominent at all. With respect to level-2, level-3 and level-4 marks on the one hand, and perception marks on the other, it is observed that the correspondences are relatively and locally manifested, but not absolutely. The fact that each sentence was read by just one speaker did not allow us to correct for speaker-dependent pronunciation. The next step is to verify the proposal on the basis of a different set of sentences from the same corpus of read-aloud sentences, as well as on the basis of a corpus which takes into a account the pronunciation of a larger group of speakers.

6. References

Baart, J. (1987). *Focus, Syntax and Accent Placement.* Dissertation, Leiden University.

Booij, G. (1995). *The Phonology of Dutch.* Oxford University Press.

Damhuis M., Boogaart T., in 't Veld C., Versteijlen M., Schelvis W., Bos L., Boves L. (1994). "Creation and analysis of the Dutch Polyphone corpus", *ICSLP 94*, Yokohama 1803 - 1803.

Dirksen, A. & H. Quené (1993). "Prosodic analysis: The next generation", in V. J. van Heuven & L. C. W. Pols (eds.) *Analysis and Synthesis of Speech*, Mouton de Gruyter, Berlin-New York, 131-144.

Dirksen, A. & L. Menert (1997). *Fluent Dutch Text-to-Speech*, Version 1.0, Fluency Speech Technology, Utrecht.

Halle, M. & J-R. Vergnaud (1987). *An essay on stress.* Cambridge, Mass.: MIT Press.

Hayes, B. (1984). The Phonology of Rhythm in English. *Linguistic Inquiry* **15**. 33-74.

Helsloot, C.J. (1995). *Metrical Prosody. A Template-and-Constraint Approach to Phonological Phrasing in Italian.* HIL Dissertation 16, HAG, Den Haag.

Kruyt, J. (1985). *Accents from Speakers to Listeners.* Dissertation , Leiden University.

Kuijpers, C., van Donselaar, M.(1998). "The Influence of Rhythmic Context on Schwa Epenthesis and Schwa Deletion in Dutch", Language and Speech **41** (**1**), 87-108.

Lea, W. (1979). "Testing linguistic stress rules with listeners' perception", in J. Wolf & D.H. Klatt (eds.) *Speech Communication Papers presented at the 97th meeting of the ASA*, New York.

Liberman, M. & A. Prince (1977). On stress and linguistic rhythm. *Linguistic Inquiry* **8**. 249-336.

Prince, A. (1983). Relating to the grid. *Linguistic Inquiry* **14**. 19-100.

Quené, H. & R. Kager (1993). "Prosodic sentence analysis without exhaustive parsing", in V. J. van Heuven & L. C. W. Pols (eds.) *Analysis and Synthesis of Speech*, Mouton de Gruyter, Berlin-New York, 115-130.

Selkirk, E. (1980). "Prosodic domains in phonology: Sanskrit revisited", in M. Aronoff and M.-L. Kean (eds.), *Juncture* (*Studia linguistica et philologica 7*). Saratoga, California: Anma Libri. 107-129.

Selkirk, E. (1986). "On Derived Domains in Sentence Phonology", *Phonology Yearbook* **3**. 371-405.

Streefkerk B. M. (1997). "Acoustical correlates of prominence: A design for research", *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, 21 131-142.

Streefkerk, B. M., Pols, L. C. W. and Ten Bosch, L. F. M. (1997) "Prominence in read aloud sentences, as marked by listeners and classified automatically", *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, 21: 101-116.

Streefkerk, B. M. & L. Pols (1998). "Prominence in read aloud Dutch sentences as marked by naive listeners" *Tagungsband KONVENS-98*, Frankfurt a.M., 201-205.

O'Shaughnessy, D. (1976). *Modelling Fundamental Frequency and its Relationship to Syntax, Semantics and Phonetics.* Dissertation MIT, Cambridge, Mass.

O'Shaughnessy, D & J. Allen (1983). "Linguistic modality effects on fundamental frequency in speech", *Journal of the Acoustic Society of America* **74/4**, p. 1155-1171.

Visch, E. (1989). *A Metrical Theory of Rhythmic Stress Phenomena.* Dordrecht: Foris.