38

Prosodic and Intonational Domains in Speech Synthesis

Erwin C. Marsi
Peter-Arno J. M. Coppen
Carlos H. M. Gussenhoven
Toni C. M. Rietveld

ABSTRACT An intonational domain corresponds to the part of an utterance spanned by one intonation contour. We lay out a theory of intonational domains that is rooted in intonational (autosegmental) phonology and prosodic phonology. We focus on restructuring—the process that joins two intonational domains together to form a single domain. We report on a perception experiment about restructuring involving synthetic speech. The results indicate that restructuring is constrained by: (1) syntactic structure, at least the distinction between a PP that is internal and a PP that is external to an NP; and (2) the length of the initial domain before restructuring. Finally, we discuss the consequences of our results for phonological theory and the intonational component in speech synthesis.

38.1   Introduction

Now that segmental synthesis has reached a certain quality, further improvement of synthetic speech is expected from improvements in the synthesis of prosody. Intonation is often considered the most salient aspect of prosodic structure, and consequently the synthesis of well-formed and contextually appropriate intonation contours has received much attention. Theoretical work on intonation, supported by both linguistic intuitions and experimental data, has accumulated over the years to constitute a field known as *intonational phonology* [Lad92]. Autosegmental descriptions of intonation, inspired by the work of [Pie80], have become available for a number of languages, including Dutch [Gus88b, Gus91]. Prosody has also been studied within the field of *prosodic phonology* [Sel84, NV86]. We feel that the synthesis of intonation can benefit from these developments in intonational and prosodic phonology. The experiment reported in this chapter is intended as a contribution to the application of autosegmental models of intonation to the synthesis of intonation.

   We use *intonational domain* to refer to each part of an utterance spanned by a single intonation contour. It is intended to be a theory-neutral expression for what is

called, for example, *intonational phrase* [NV86, Sel84], *intermediate phrase and intonation phrase* [BP86], *tone domain* [Lad86], or *association domain* [Gus88a]. In what follows we concentrate on intonational domains and abstract away from other aspects of intonation, such as the particular shape of the intonation contour (*tune*) or the prominence of pitch accents. In the first part of this chapter we lay out our theory of intonational domains, which is based on linguistic intuitions and practical experience with speech synthesis, and compare it to the conventional conception of intonational domains in prosodic phonology. Subsequently we focus on the problem of the *restructuring* of intonational domains—the phenomenon that under certain conditions adjacent intonational domains are joined together to form a single such domain. The problem of restructuring directly motivated the experimental work that is described in the second part of this chapter. This work concerned the effect of syntactic structure and length on the distribution of intonational domains.
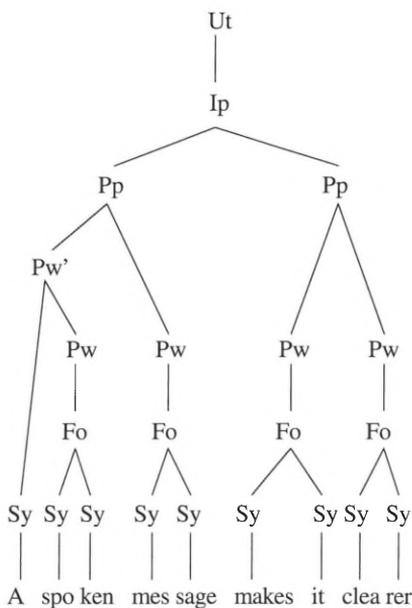
38.2   A Theory of Intonational Domains

38.2.1   Intonational Domains and the Prosodic Hierarchy

The status of the intonation domain in phonological structure has been discussed in a number of recent publications. When the theory of prosodic phonology was first introduced, phonologists assumed that the intonation contour corresponded to a constituent called the *intonational phrase* (Ip), which took its place in a hierarchy of other prosodic constituents [Sel84, NV86]. Prosodic constituents are defined by means of mapping rules that derive the prosodic constituency of an utterance from its morphosyntactic structure. An utterance such as "A spoken message makes it clearer" might look like (1), where the Ip is slotted in between the *utterance* (Ut) and the *phonological phrase* (Pp), below which are the *phonological word* (Pw), the *foot* (Fo) and the *syllable* (Sy).[1]

The general motivation for prosodic phonology is that prosodic constituents appear to figure in the structural description of segmental phonological rules, as amply shown in [NV86]. It has also been assumed that these constituents define the rhythmic structure of the utterance, either directly or in more recent accounts, after translation into a metrical grid [NV89]. In the case of the Ip, there appears to be a problem when the criterion for the constituency provided by the intonation contour is confronted with the criteria provided by pausal and segmental phenomena. The difficulty is that the intonation contours that are appealed to in the definition of the Ip may have internal prosodic breaks that in other cases typically occur at the boundaries of the Ip. As an example, consider the utterances given in (2). Example (2a), where "John" is the indirect object of "tell," is a single intonation contour that
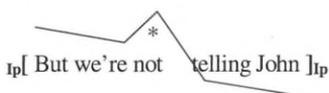
---

[1]The specific way in which function words have been attached in (1) is not relevant to the subsequent discussion.
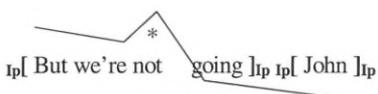
(1)

```
                              Ut
                              |
                              Ip
                        _____/_____
                       /              \
                      Pp              Pp
                    /\               /\
                 Pw'  \             /  \
                /\     \           /    \
              Pw   Pw   Pw        Pw    Pw
               |    |    |         |     |
              Fo   Fo   Fo        Fo    Fo
              /\   /\   /\        /\    /\
           Sy Sy Sy Sy Sy        Sy   Sy Sy Sy
            |  |  |  |  |         |    |  |  |
            A spo ken mes sage  makes  it clea rer
```

(2)

a.

Ip[But we're not   telling John]Ip
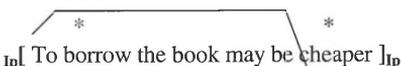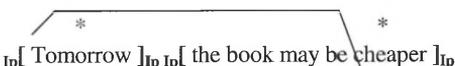
b.

Ip[But we're not   going]Ip Ip[John]Ip

corresponds to a single Ip. Example (2b), where "John" is a vocative tag, has the same contour as (2a) but is analyzed as containing two Ip's. Primarily because of the pause that separates them from the preceding clause, vocatives are interpreted as separate Ip's. But because final vocatives are unaccented, the intonation contour chosen for the accent on "not" (the only accent in the Utterance) necessarily spans two Ip's (cf. [Tri59]). Thus, in (2b), a single intonation contour is mapped onto more than one Ip.

In accented Ip's, too, this conflict may arise. For example, in (3b), the preposed adverbial "Tomorrow" occurs in exactly the same intonational contour as does the verb "to borrow" in (3a). That is, on the basis of the intonational criterion, both (3a) and (3b) ought to be single Ip's. However, on the basis of the pausal and segmental

(3)

a.

$_\text{Ip}$[To borrow the book may be cheaper]$_\text{Ip}$

b. $_\text{Ip}$[Tomorrow]$_\text{Ip}$ $_\text{Ip}$[the book may be cheaper]$_\text{Ip}$

phenomena, they are distinct. There is a rhythmic break following "Tomorrow" in (3b) which is absent after "To borrow" in (3a)[2]

38.2.2 Association Domains

A response to this problem is given in [Gus88a]. Gussenhoven argues that the solution should be based on the recognition that the intonational domain of a pitch accent cannot consistently be identified with any one constituent in the prosodic hierarchy. The intonational domain of a pitch accent is determined primarily by the location of any following accent, and only secondarily by constituent structure. Therefore, he assumes that we need an independent constituent over which a single intonation contour spreads: the *association domain* (AD). The AD is determined by two factors: the pitch accent distribution and the prosodic constituency. The properties of ADs and their relation to accent distribution in combination with prosodic constituency can be stated as a sequence of constraints.

Constraints on Association Domains:

   A Every pitch accent has its own unique AD and every AD belongs to a unique pitch accent.

   B The AD of an accent ends with the *highest* prosodic constituent that dominates the accent but does *not* dominate a following accent.

   C Every AD has a maximum size without overlapping other ADs.

As a consequence of constraint A, ADs and pitch accents are in a one-to-one relation. The situation of an AD without an accent or an accent without an AD does not occur[3] Constraint B expresses the important fact that ADs depend on the number and locations of accents in the utterance. Furthermore, this constraint guarantees that the end of every AD is aligned with the end of a prosodic constituent. Notice that divorcing intonational domains from prosodic constituency does not imply that intonational domains do not respect the boundaries of prosodic constituents. Rather, the consequence of our view is that instead of there being one particular

---

[2]Cf. [GR92b] for segmental phenomena that also indicate a distinction.

[3]There is one exception: ADs without a pitch accent may occur for reporting clauses. In such cases a rule called *tone copy* provides a pitch contour to the empty AD (cf. [Tri59].

(4)
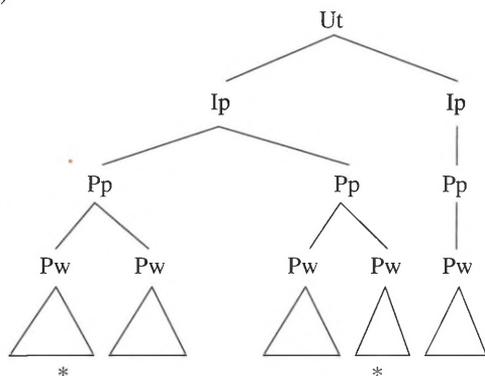
$_{AD}[$ A spoken message $]_{AD}$ $_{AD}[$ makes it clearer, he said $]_{AD}$
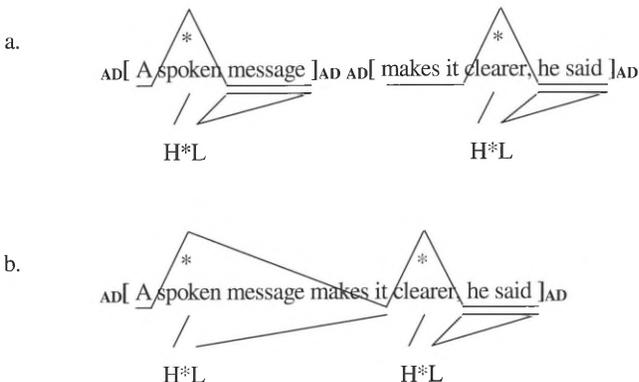
prosodic constituent that can be identified with the intonational domain of a pitch accent, any one of a number of constituents from the foot onward can define the rightmost boundary of a domain. Finally, constraint C implies that the left-hand boundary of an AD will be positioned to the left as far as possible, coinciding either with the end of a previous AD or the beginning of the utterance.

In (4), an example is given that shows how we arrive at a unique distribution of ADs by application of the constraints. Consider the pitch accent on "spoken." As a consequence of constraint A, it must have its own unique AD. Constraint B says that the end of this AD coincides with the highest prosodic constituent that dominates the accent, but does not dominate the next accent (on "clearer"). This can only be the first Pp. It cannot be the Pw that dominates "A spoken" because this is not the highest constituent. Neither can it be the first Ip or any higher constituent because these also dominate the next accent. The AD starts at the beginning of the utterance by virtue of constraint C. Next, consider the accent on "clearer." As the Ut is the highest constituent that dominates it and there are no accents following, the end of its AD coincides with the end of the Ut. To maximize the size of the AD without overlap, the AD starts where the previous one ends. Now, all constraints are satisfied, and every accent has its own AD. The shape of the actual intonation contour depends on the type of pitch accent with which the accents will be realized.

38.2.3   AD-Restructuring

It has so far been assumed that there is a one-to-one relationship between pitch accents and ADs. However, this situation is characteristic of slow and emphatic speech only. The intonation of ordinary natural speech shows that often several pitch accents actually share a common intonational domain. We assume there is an optional cyclic process called *AD-restructuring* that joins two adjacent ADs to form a single AD. The intonational consequence of restructuring can be illustrated with the help of the representations in (5). The "*" marks an accented syllable. The bottom line represents the tonal string that consists of pitch accents. The two strings

(5)

a.

AD[A spoken message]AD AD[makes it clearer he said]AD

H*L   H*L

b.

AD[A spoken message makes it clearer he said]AD

H*L   H*L

are aligned by means of association lines. The starred tone of the pitch accent, H*
here, goes to the accented syllable, "spo" in this case. The second tone of the
pitch accent, L here, *spreads*. This means that the pitch in the stretch of speech
until the right-hand AD boundary is determined by the L. Spreading of a tone
is indicated by a triangle.[4] In (5b), the two ADs have been restructured, causing
the first AD to lose its right-hand boundary. As a result, the final tone of the first
pitch accent moves to the syllable just before the second accented syllable.[5] After
interpolation between the targets corresponding to the tones, the intonation contour
has a different shape. In addition to the intonational consequences of restructuring,
there are also durational consequences. [GR92b] provides experimental evidence
for the claim that an AD boundary enhances *preboundary lengthening*, that is,
a syllable before an AD boundary is longer than a syllable that is not AD-final.
Furthermore, a right-hand AD boundary is often followed by a pause, which may
also disappear after restructuring (see also [SC94]).

### 38.2.4   Constraints on AD-Restructuring

One of the problems concerning restructuring can be illustrated with the help of
example (6). Part (a) shows the intonation contour before any restructuring took
place. Restructuring of the first two ADs results in part (b), which is a perfectly
normal intonation. Conversely, restructuring of the last two ADs gives rise to part

---

[4]The unaccented stretches preceding the first accent in an AD, i.e., "a" in the first
and "makes" in the second, will typically have low or mid-pitch. We account for this by
assuming an AD-initial *boundary tone* that spreads up to the first pitch accent. Also, we
assume utterance-final boundary tones: L% or H%. We have omitted the boundary tones in
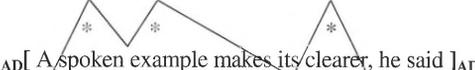the examples.

[5]In our phonological model the symbols on the tone string are associated with syllables
on the segmental string. The phonetic model has the task of determining the exact alignment
within syllables (cf. [GR92a]).

(6)

a.

_{AD}[A spoken]_{AD} _{AD}[example]_{AD} _{AD}[makes its clearer, he said]_{AD}

b.
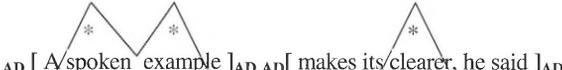
_{AD} [A spoken example]_{AD} _{AD}[makes its clearer, he said]_{AD}

c.
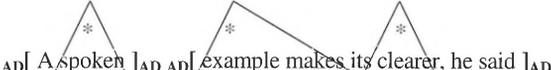
_{AD}[A spoken]_{AD} _{AD}[example makes its clearer, he said]_{AD}

d.

_{AD}[A spoken example makes its clearer, he said]_{AD}

(c), which is a very unnatural, if not ill-formed, intonation. However, restructuring of all three ADs results in a normal intonation again (part (d)). On the basis of examples such as this, [GR92b] proposed the following rule: *restructure two adjacent ADs, giving precedence to ADs that are separated by a lower-ranking prosodic boundary.* Notice that the first two ADs are separated by a Pw boundary whereas the last two ADs are separated by a PP boundary (cf. (4)). Thus, in the case of example (6), the proposed rule correctly excludes the restructuring in part (c), because the one in part (b) should have precedence.

Nevertheless, there are still many situations in which this rule offers no solution, simply because two or more pairs of adjacent ADs are separated by a prosodic boundary of the same strength. Examples like (7) suggest that the order of restructuring is nevertheless meaningful. This is an example of a well-known ambiguity caused by two possible interpretations of the PP (prepositional phrase) "with a telescope." The restructuring in (7a) favors the interpretation of the PP as a modifier of the verb "saw," that is, the policeman saw the criminal by means of a telescope. The restructuring in (7b), on the other hand, favors the interpretation of the PP as a modifier of the noun "criminal," that is, the policeman saw the criminal who possessed a telescope. Examples such as those in (7) suggest that restructuring is somehow related to the syntactic structure of an utterance. The experimental work described in the next section tries to establish this relation in the case of nonambiguous sentences.

(7)

a.   *   *   *
    AD[The policeman saw the criminal]AD AD[with the telescope]AD

b.   *   *   *
    AD[The policeman]AD AD[saw the criminal with the telescope]AD

38.3   Restructuring Intonational Domains: An Experiment

38.3.1   Hypotheses

In section 38.2, we pointed out a problem with AD-restructuring. We concluded that prosodic constituency does not sufficiently constrain the restructuring of ADs. We hypothesized, on the basis of ambiguous sentences, that syntactic structure might govern the restructuring of ADs. In addition, the length of the first AD to be restructured appears to influence the probability of restructuring. This led us to an experiment in which we investigated the dependence of restructuring on two factors: (1) the syntactic structure of an utterance; (2) the length of the first AD.

We selected two syntactic structures whose analysis is relatively uncontroversial, and moreover seem intuitively likely to influence restructuring in opposite ways. In fact, they are the syntactic structures that correspond to each of the two interpretations of the ambiguous example in (7). In the first structure the PP serves as a noun modifier. In terms of tree structures, this means that the PP is attached somewhere internal to an NP (noun phrase). In the second structure, by contrast, the PP serves as either a predicate modifier or a separate argument. This means, again in terms of tree structures, that the PP is not attached internal to an NP.

In view of the fact that intonation can contribute to linguistic processing, that is, the recovering of the structural aspects of the message, it seems reasonable to consider intonational domains as one of the cues to the attachment of a PP. For syntactically ambiguous utterances such as those shown in (7), the intonational domains might in fact be the only cue to the right interpretation. But even if the sentence in question is not ambiguous, intonational domains could still form an auxiliary cue. Moreover, incompatible intonational domains would clash with syntactic cues, thereby complicating the processing. In the case of an external PP, for example, we expect an intonational boundary (i.e., the start of a new AD in terms of our theory) to coincide with the beginning of the PP. Conversely, we expect no intonational boundary in the case of an external PP, that is, we expect AD-restructuring. Hypothesis 1 reflects these expectations concerning the relation between the attachment of the PP on the one hand and restructuring on the other.

Hypothesis 1: Restructuring based on syntactic structure

    a If a PP is internal to an NP, restructuring will force them in the same AD.

b If a PP is external to an NP, no restructuring will occur.

Clearly, there are limits to the acceptable length of an intonational domain. In cases where restructuring would give rise to unacceptably long ADs, restructuring is not allowed. Conversely, restructuring of a relatively short AD could take place even if this would cause an NP and an external PP to be in the same AD. In other words, we expected that the factor of the length of the initial AD would sometimes override the factor of syntactic structure. This expectation is reflected in hypothesis 2: The conditions that length places on restructuring can overrule restructuring based on syntactic structure.

Hypothesis 2: Restructuring based on length

a If the length of an AD is "relatively long," restructuring will be blocked (overruling hypothesis 1a).

b If the length of an AD is "relatively short," restructuring will be forced (overruling hypothesis 1b).

The experiment was meant to obtain evidence for the two hypotheses as well as more specific indications of what "relatively long" and "relatively short" might mean.

We decided to use synthesized speech in this experiment. This allowed us to manipulate the intonational aspects only. Therefore, we could produce several versions of the same utterance that differed only with respect to their intonation contours and intonational domains. The decision to use synthesis instead of resynthesis was motivated by our aim to improve the intonational component of our speech synthesizer.

38.3.2 *Material*

We devised four sets of sentences with PPs internal to NPs, and four sets with PPs external to NPs. Each set contained four sentences that differed only in the length of the stretch from the start of the sentence to the beginning of the PP, measured in terms of the number of syllables. There were four corresponding ranges: [5–7] (mean value 5.9), [9–12] (mean value 10.4), [13–17] (mean value 15.1), and [17–20] (mean value 19.1) syllables. The stretch from the beginning of the PP to the end of the sentence was kept relatively short ([6–8] syllables, mean value 7). Examples of these sets are given in (8a) and (9a); (8b) and (9b) provide a gloss.

None of these sentences were ambiguous for humans. We accented all the words except verbs,[6] pronouns, and complementizers. Although in Dutch an optional rhythmic adjustment rule allows for the deletion of some accents, we decided to distribute the accents in a conservative but clearly acceptable way. For each sentence, we derived the corresponding strictly layered prosodic structure by means

---

[6]Verbs in Dutch can remain unaccented even when they are included in the focus of the sentence. The same is true for English and German [Sel95].

(8)

a.1   Tenzij hij zijn gedachten
       over het schaakspel bedoelde.

a.2   Tenzij hij zijn filosofische gedachten
       over het schaakspel bedoelde.

a.3   Tenzij hij zijn onorthodoxe filosofische gedachten
       over het schaakspel bedoelde.

a.4   Tenzij hij zijn bijzonder onorthodoxe filosofische gedachten
       over het schaakspel bedoelde.

b.   Tenzij hij zijn bijzonder unorthodoxe filosofische   gedachten
       unless he his very   unorthodox philosophical thoughts
       over het schaakspel   bedoelde.
       about the chess-game meant
       "Unless he meant his very peculiar philosophical thoughts
       about the game of chess."

c. $_S$(tenzij hij zijn $_{NP}$($_{NP}$(zijn ... gedachten)$_{NP}$ $_{PP}$(over het schaakspel)$_{PP}$)$_{NP}$
       bedoelde)$_S$

d. $_{Ut}$($_{Ip}$($_{Pp}$(tenzij hij zijn ... gedachten)$_{Pp}$ $_{Pp}$(over het schaakspel)$_{Pp}$
       $_{Pp}$(bedoelde)$_{Pp}$)$_{Ip}$)$_{Ut}$

e.1

       $_{AD}$[tenzij hij zijn ... gedachten]$_{AD}$ $_{AD}$[over het schaakspel bedoelde]$_{AD}$

       L   H*LH   L   H*L   L%

e.2

       $_{AD}$[tenzij hij zijn ... gedachten over het schaakspel bedoelde]$_{AD}$

       L   H*L   H*L   L%
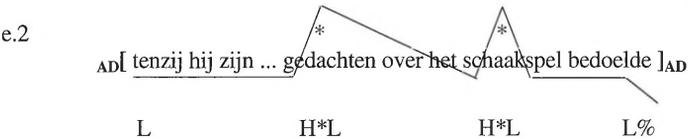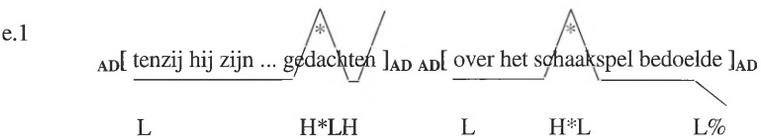
of the mapping rules described in [NV86]. (8d) and (9d) exemplify the prosodic structures derived from the respective syntactic structures in (8c) and (9c). Notice that although their syntactic structures differ, their prosodic structures are identical; this is exactly the situation we wanted to investigate. Next, we determined for every accent its AD on the basis of accent distribution and prosodic structure. We assumed

that all ADs separated by only a Pw boundary were obligatorily restructured. The net effect was that sentences consisted of two ADs, with the beginning of the second AD corresponding to the beginning of the PP. We produced two utterances out of each sentence by means of a rule-based allophone speech synthesizer for Dutch [KRGE94]. The first version contained two ADs, whereas the second restructured version contained only a single AD. The contrast in the shape of the intonation contours is exemplified in example (8e.1) versus (8e.2). The final pitch accent of the initial AD in the versions containing two ADs was realized as H*LH in order to obtain an H boundary tone, whereas all other pitch accents were realized as H*L. These choices of pitch accents are conservative in the sense that the resultant tunes are maximally neutral. In pilot experiments, it had become clear that our naive raters experienced difficulties when judging stylized intonation contours. We decided to increase the naturalness of the synthetic utterances by marking the nonfinal right-hand AD boundaries by a lengthening of preboundary syllable (10–30 ms) and by adding a small postboundary pause (35 ms).[7] We also added an accentual and phrasal *downstep* of 0.9 to each of the utterances, which caused the subsequent F_0 peaks to be lower than the preceding peak by a constant factor [vGR92].

38.3.3   Method

The restructured and nonrestructured versions of each sentence were compared by 25 naive raters. Raters could read the orthographic version of the sentence on a screen and could listen to both versions as many times as they wished. In the case of naive raters, it is, of course, impossible to ask directly which intonation contour is the more appropriate one, given the syntactic structure of the sentence. Instead, we asked them to judge which of the two versions was more adequate, given the meaning of the sentence. The total of 32 pairs of utterances were presented in random order. Raters recorded their ratings on a 7-point scale that ran from "first version much better" via "equally good" to "second version much better." In order to focus their attention on the phenomenon that we wanted them to judge, we started every session with four familiarization pairs that contained an idiomatic expression or proverb. In expressions such as these, the choice of intonational domains is fixed by convention. One version of every training pair had the conventional AD boundary, and the other version contained a very unlikely AD boundary inside the idiomatic expression. Raters received feedback about the correctness of their rating during these familiarization utterances.

---

[7]It is generally accepted that these amounts of lengthening and pausing are by themselves not sufficient to produce a subjective break.

(9)

a.1   Tenzij hij zijn gedachten

in een vakblad publiceert.

a.2   Tenzij hij zijn filosofische gedachten

in een vakblad publiceert.

a.3   Tenzij hij zijn onorthodoxe filosofische gedachten

in een vakblad publiceert.

a.4   Tenzij hij zijn bijzonder onorthodoxe filosofische gedachten

in een vakblad publiceert.

b.   Tenzij hij zijn bijzonder unorthodoxe filosofische   gedachten

unless he his very   unorthodox philosophical thoughts

in een vakblad   publiceert.

in a   professional-journal publishes

"Unless he publishes his very peculiar philosophical thoughts

in a professional journal."

c. $_S$(tenzij hij zijn $_{NP}$(zijn ... gedachten)$_{NP}$ $_{PP}$(in een vakblad)$_{PP}$ publiceert)$_S$

d. $_{Ut}$($_{Ip}$($_{Pp}$(tenzij hij zijn ... gedachten)$_{Pp}$ $_{Pp}$(in een vakblad)$_{Pp}$

$_{Pp}$(publiceert)$_{Pp}$)$_{Ip}$)$_{Ut}$

38.3.4   Results

An analysis of variance was carried out on the data, with three within-subject factors: (1) syntactic structure, (2) length of the initial AD, and (3) sentence set. First of all, the analysis revealed no significant interaction among the factors syntactic structure and length of initial AD. Furthermore, it showed that the syntactic structure is a significant factor ($F_{1,24} = 13.79$, $p < 0.001$). This implies that restructuring is indeed sensitive to whether attachment of a PP is internal or external to an NP. The length of the initial AD is also a significant factor ($F_{3,72} = 7.25$, $p < 0.001$, Huynh-Feldt corrected). As expected, restructuring is sensitive to the length of the initial AD as well.

Figure 38.1 shows the combined effect of the factors syntactic structure and length of initial AD. Apparently, restructuring is less acceptable for utterances with an external PP than for utterances with an internal PP. Likewise, the acceptability of restructuring decreases when the length of the initial AD increases. z-ratios (one-tailed) were used to determine whether some combinations of the two factors made the preference scores significantly larger or smaller than zero ($p < 0.05$),
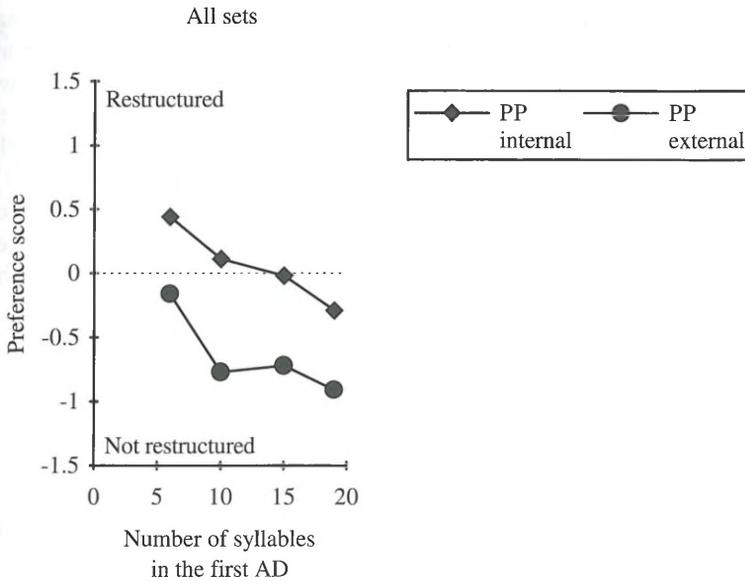
All sets

FIGURE 38.1. Mean preference scores over all four sets.

as this would imply a significant preference for either restructuring or a boundary. Restructuring is significantly preferred for those utterances with an internal PP and an initial AD of six or fewer syllables. On the other hand, a boundary instead of restructuring is significantly preferred in the case of an internal PP and an initial AD of 19 or more syllables. Finally, utterances with an external PP and a initial AD of 10 or more syllables are significantly preferred without restructuring. The only significant interaction, according to the analysis of variance, occurred among the factors length and sentence set.

These results support the claim that both the syntactic structure of an utterance and the length of its ADs are relevant to AD-restructuring. However, they are incompatible with hypothesis 1a, which states that if a PP is internal to an NP, restructuring will cause them to be in the same AD. Notice that it would be supported if we restricted ourselves to only those utterances with the shortest initial AD. Here, the preferred restructuring is more appropriately explained by referring to the length of the initial AD, as in hypothesis 2. The results do, however, support hypothesis 1b, which states that if a PP is external to an NP no restructuring will occur. The indecisive behavior of those utterances with the shortest initial AD is, again, more appropriately explained by the factor length.

For restructuring based on length we arrive at a similar conclusion. Hypothesis 2a states that restructuring will be blocked if the length of the initial AD is "relatively long." If we substitute the value "19 or more syllables" for "relatively long," this hypothesis is in fact compatible with the data (cf. the first point from the right for PP internal in figure 38.1). The expectation that restructuring will be forced if the

length of an AD is "relatively short," as formulated in hypothesis 2b, can be neither rejected nor supported. The mean preference score for utterances with an external PP in combination with the shortest initial AD does not significantly differ from zero (cf. the first point from the left for PP external in figure 38.1). Although we cannot decide on hypothesis 2b in general, it is supported if we restrict ourselves to utterances with an internal PP and substitute "six or fewer syllables" for "relatively short" (cf. the first point from the left in the graph for PP internal in figure 38.1). In addition, we can remark that a boundary is no longer significantly preferred for utterances with an external PP and the smallest initial AD, as opposed to utterances with an external PP and a larger initial AD (cf. the first point from the left in the graph for PP external in figure 38.1).

38.4   Discussion

38.4.1   Effect Size

Although some of the results turned out to be statistically significant, they do not indicate a great sensitivity of the raters to any of the examined factors when compared to the full range of the scale. A possible explanation is that most raters considered the task to be rather difficult. An examination of the scores reveals that raters were somewhat erratic. For example, 16 out of 25 raters recorded at least once a difference of 4 points between pairs of utterances that differed in only one additional word. However, in support of our results, we can mention that they agree with other experimental results. Sanderman and Collier [SC94] report that both the length of the initial intonational domain and attachment of the PP influence the *perceptual boundary strength* (PBS).

Another point is the relation to speech rate. It is generally assumed that increasing the speech rate will reduce the number of intonational boundaries; see [Cas94] for experimental evidence. In our stimuli, we aimed to reproduce a normal speech rate such as that used in news broadcasting. We can only speculate about the effects on our results of an increased speech rate. Perhaps this would shift up all points in figure 38.1 and confirm hypothesis 1b.
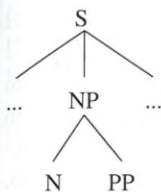
Finally, we have seen that the critical AD lengths for restructuring are somewhere around 5 and 15 syllables. It would be interesting to see if these figures can be confirmed in a follow-up experiment.

38.4.2   Consequences for Our Theory of Intonational Domains
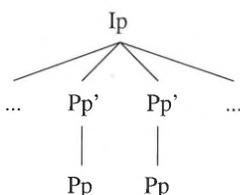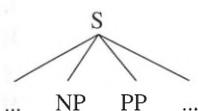
One of the reasons for assuming prosodic structure is that syntactic structure is too informative for the purposes of phonology. Prosodic phonology interfaces between syntax and phonology. In the process mapping the syntactic constituents to phonological constituents the amount of structural information is reduced. Recursive structure, for example, is one of the syntactic aspects that will be removed. The claim that prosodic structure contains no recursive constituents is known as the

(10)

Syntactic structure: Prosodic structure:

```
       S                                                      Ip
     /⌃\                                                    /⌃\
 ...  NP  ...                    ⟶                    ...  Pp'  ...
     /\                                                    /\
    N   PP                                               Pp   Pp
```

```
       S                                                      Ip
     /⌃\                                                    /⌃\
 ... NP  PP ...                  ⟶                    ...  Pp'  Pp'  ...
                                                           |    |
                                                          Pp   Pp
```
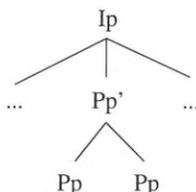
strict layer hypothesis. Exactly this claim is responsible for the problem that the results pose for our theory of intonational domains. Restructuring apparently depends on the way a PP is attached, internal or external, to a NP, but this information is lost during the translation from syntactic to prosodic structure. Both syntactic structures map onto the same prosodic structure (cf. example (8cd) and (9cd)). Consequently, the constraints on restructuring cannot be adequately expressed.

A possible solution is to augment the prosodic hierarchy with a recursive phonological phrase, let's say, a Pp'. The mapping rules would have to be adapted to give the result as in (10), which would enable us to express the difference in restructuring. Deleting the AD boundary between a noun and an internal PP, which coincides with a Pp boundary, has priority over the deletion of the AD boundary between an NP and external PP, which coincides with a Pp' boundary. Of course, this solution amounts to a rejection of the strict layer hypothesis. In fact, other research exists that suggests that the restrictions the strict layer hypothesis imposes on the prosodic structure are actually too strong [GG83, Lad86, dS94, p. 2046].

38.4.3   Consequences for Speech Synthesis

The present results allow us to improve our intonational domains in speech synthesis. In text generation, the information about the attachment of PPs as well as the length of the ADs is readily available. Real ambiguous sentences (to humans), in which the intonational domain is the only cue left to the right interpretation, are quite rare. On the other hand, PPs themselves are very frequent. In the latter case,

adequate intonational domains will most likely contribute to processing ease and perceived naturalness.

38.5   Conclusion

Intonational domains cannot be consistently identified with the intonational phrase of the prosodic hierarchy. We need an independent domain, called the association domain, which is determined primarily by the pitch accent distribution and secondarily by the prosodic constituency. In principle, each accent has its own AD, but the process called AD-restructuring can join two ADs to form a single one. Restructuring seems to depend on the syntactic structure: It becomes more likely when an AD boundary corresponds to the beginning of a PP that modifies a noun. In addition, restructuring seems to depend on the length of the ADs involved: It is negatively related to the length of the first AD. The constraints on restructuring cannot be adequately expressed in terms of conventional prosodic constituency, which respects the strict layer hypothesis.

REFERENCES

[BP86]   M. Beckman and J. Pierrehumbert. Intonational structure in English and Japanese. *Phonology Yearbook* 3:255–309, 1986.

[Cas94]   J. Caspers. *Pitch Movements Under Time Pressure: Effects of Speech Rate on the Melodic Marking of Accents and Boundaries in Dutch.* (HIL dissertations 10) Holland Academic Graphics, Den Haag, 1994.

[dS94]   J. R. de Pijper and A. A. Sanderman. On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *JASA* 96(4):2037–2047, 1994.

[GG83]   J. P. Gee and F. Grosjean. Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology* 15:411–458, 1983.

[GR92a]   C. Gussenhoven and T. Rietveld. A target interpolation model for the intonation of Dutch. In *Proceedings of ICSLP 92*, vol. 2, 1235–1238, 1992.

[GR92b]   C. Gussenhoven and T. Rietveld. Intonation contours, prosodic structure and preboundary lengthening. *J. Phonetics* 20:283–303, 1992.

[Gus88a]   C. Gussenhoven. Intonational phrasing and the prosodic hierarchy. In *Phonologica 1988*, W. U. Dressler et al., eds. Cambridge University Press, Cambridge, 89–99, 1988.

[Gus88b]   C. Gussenhoven. Adequacy in intonation analysis: The case of Dutch. In *Autosegmental Studies on Pitch Accent*, H. v. d. Hulst and N. Smith, eds. Foris Publications, Dordrechet, 95–121, 1988.

[Gus91]   C. Gussenhoven. Tone segments in the intonation of Dutch. In *The Berkeley Conference on Dutch Linguistics 1989*, T. F. Shannon and J. . Snapper, eds. University Press of America, Lanham, MD, 139–155, 1991.

[KRGE94] J. Kerkhoff, T. Rietveld, C. Gussenhoven, and L. Elich. NIROS: The Nijmegen Interactive Rule Oriented Speech Synthesis System. Proceedings, Department of Language and Speech 18, University of Nijmegen, 107–119, 1994.

[Lad86]   D. R. Ladd. Intonational phrasing: The case for recursive prosodic structure. *Phonology Yearbook* 3:311–340, 1986.

[Lad92]   D. R. Ladd. An introduction to intonational phonology. In *Papers in Laboratory Phonology II*, G. J. Docherty and D. R. Ladd, eds. Cambridge University Press, Cambridge, 321–334, 1992.

[NV86]   M. Nespor and I. Vogel. *Prosodic Phonology*. Foris, Dordrecht, 1986.

[NV89]   M. Nespor and I. Vogel. On clashes and lapses. *Phonology* 6:69–116, 1989.

[Pie80]   J. Pierrehumbert. *The Phonology and Phonetics of English Intonation*. Garland Press, New York, 1980.

[SC94]   A. A. Sanderman and R. Collier. Prosodic phrasing at the sentence level. In *Festschrift for K. Harris of Physics, Modern Acoustics and Signal Processing Series*, American Institute of Physics, 1994.

[Sel84]   E. Selkirk. *Phonology and Syntax*. MIT Press, Cambridge, MA, 1984.

[Sel95]   E. Selkirk. Sentence prosody: Intonation, stress and phrasing. In *Handbook of Phonological Theory*, J. Goldsmith, ed. Blackwell, 550–569, 1995.

[Tri59]   J. L. M. Trim. Major and minor tone groups in English, *Le Maître Phonétique* 112:26–29, 1959.

[vGR92]   R. van den Berg, C. Gussenhoven, and T. Rietveld. Downstep in Dutch: Implications for a model. In *Papers in laboratory phonology II*, G. J. Docherty and D. R. Ladd, eds. Cambridge University Press, Cambridge, 255–310, 1992.