6

What is the smallest prosodic domain?

VINCENT J. VAN HEUVEN

It is widely held that the syllable is the smallest prosodic domain. Notions such as stress, accent, and preboundary lengthening are typically defined as properties of an entire syllable. This paper considers another possibility: that single segments may also function as prosodic domains below the syllable. Within a syllable, if any of the segments is placed in narrow (contrastive) focus, is it prosodically marked by the speaker, e.g. by melodic and/or temporal means? If so, then accent must be a property of the segment, not the syllable, and each segment must be a prosodic domain. And if that is the case, then the question arises which of the segments is the head of the larger prosodic domain, the syllable. These questions can be addressed through acoustic and perceptual studies.

6.1 Theoretical considerations

6.1.1 Integrative focus, narrow focus, and accent position

Accent is defined here as prosodic prominence of a syllable (or part thereof, see below) brought about mainly by melodic means (cf. Bolinger 1958). In Dutch, for instance, it is a sufficient condition for the perception of accent that one of four different fast pitch movements is executed in an appropriate position within the syllable (cf. 't Hart *et al.* 1990).

The function of a (pitch) accent is to place a linguistic unit in focus, i.e. present the unit as expressing important information to the listener (cf. Ladd 1980; Gussenhoven 1984; Baart 1987; Nooteboom and Kruyt 1987). For example, an appropriate answer to question (1a) would be (1b):

(1) a. WHO wrote that NOVEL?

b. The DEAN of our FACULTY wrote that silly book.

In the answer both *The dean* and *of our faculty* are presented in focus by pitch accents (indicated by capitalized stressed syllables). The second part of the answer (*wrote that silly book*) contains no accent(s) and is therefore out of focus, since it is a mere repetition of material mentioned earlier in the question.

Although it would be possible to mark every (content) word in a larger focus domain by a separate pitch accent, this is not normally done. Speakers typically present several words in focus together in a coherent word group by marking only one word within the larger constituent with a pitch accent. This word is called the "exponent" (Fuchs 1984) or "prosodic head" of the constituent. Consequently, (1c) with an accent only on *faculty* would be an alternative answer to question (1a) expressing essentially the same focus distribution, i.e. presenting the entire constituent *The dean of our faculty* in focus:

(1) c. The dean of our FAculty wrote that silly book.

When a word such as *dean* in (1c) has no pitch accent, it may therefore either be out of focus (i.e. presented as less important to the listener), or be in focus as part of a larger constituent with an accent on the prosodic head elsewhere in the constituent. More generally, an accent on the prosodic head of a larger constituent is ambiguous, signifying that either the entire constituent is meant to be in focus, or only the head. Consequently, we would expect (1c) to be an appropriate answer to both questions (1a) and – with implicit negation – (1d):

(1) d. Did the dean of your CHURCH write that novel?
c. (No,) the dean of our FAculty wrote that silly book.

If a pitch accent occurs on a word other than the prosodic head of the larger domain, all other words in the larger domain are out of focus, and the accented word is presented with narrow focus, typically expressing a contrast. For instance, (1f) can only express a contrast with another faculty official, as in (1e):

(1) e. Did the SEcretary of your faculty write that novel?
f. (No,) The DEAN of our faculty wrote that silly book.

To sum up, accenting the head of a prosodic constituent yields an ambiguous focus distribution: it may signal either integrative focus on the entire domain, or narrow focus (with implicit contrast) on the accented unit only.

6.1.2 Lexical stress and pitch accent

So far we have described accent as if it were a property of an entire word. However, we can generalize the mechanism of integrative focus to situations at the word level if we define the (lexically) stressed syllable as the exponent of the word domain. Clearly, an accent on the lexically stressed syllable suffices to mark the entire word for focus, as in (2a):

(2) a. I said DIGEST, not EAT.

Here the entire word *digest* is contrasted with *eat*; yet, one accent on the lexically stressed second syllable is needed to put both the stressed and the unstressed syllable in focus.

This also accounts for the fact that accent on the lexically stressed syllable is ambiguous, since it may also express narrow focus on just the lexically stressed syllable, as in (2b), in which the realization of *digest* is indiscriminable from that in (2a); for Dutch data bearing this out see Sluijter (1992).

(2) b. I said DIGEST, not DIVERT.

Here only the lexically stressed (final) syllables of *digest* and *divert* are in focus, since the identical initial syllables *di-* are not contrasted.

It is, of course, quite possible to place the accent on a syllable that does not bear the lexical stress (Bolinger 1961). Accent on a lexically nonstressed syllable would then be a case of narrow focus, expressing a contrast below the word at the level of the syllable, as in (2c):

(2) c. I said DIgest, not SUGgest.

Here the contrast is made only for the initial lexically nonstressed syllables *di* versus *sug*, whilst the lexically stressed second syllables *gest* are identical and therefore out of focus. That only the nonstressed syllables are in focus is clear from the incorrect expressions (2d–e) in which the accent is not on the exponent of the word, even though entire words rather than individual syllables are contrasted:

(2) d. *I said DIgest, not EAT.
e. *I said SUGgest, not CLAIM.

We therefore claim that accent is a property of a syllable rather than of an entire word. Whole words are presented in focus by widening the scope of the accent on the vowel in the lexically stressed syllable.

6.1.3 Accent as a segmental property

The question that I wish to address is if, by an extension of the argument above, the focus domain can be narrowed further to a subsyllable level: for instance, to the level of the segment.

Segments within a syllable are hierarchically organized: the vocalic nucleus is more basic to the syllable than the other, consonant-like elements. A single vowel may quite well constitute a syllable on its own; even single-vowel words do occur (English: *eye*, *a*; Dutch *u* [y] "you," *ei* [ei] "egg," *ui* [ʌy] "onion"). Consonantal elements, on the other hand, can often be omitted without yielding an illegal structure, and are generally incapable of constituting syllables by themselves. Clearly, then, if the segments within the syllable are hierarchically ordered, the vowel should be the head or exponent, and the consonants the satellites.

Generally, accent is defined as a property of a constituent of at least the size of a syllable. The notable exception would be Chomsky and Halle (1968) who proposed that vowels (rather than consonants) be marked for stress (stress and accent were not differentiated). Although this proposal was primarily motivated by the circumstance that phonological theory at the time did not incorporate syllables, we have taken our cue from it: we shall work from the assumption that individual segments can be given narrow focus, and hence can be marked by an accent. Moreover, we claim that the vowel is the exponent of the syllable, so that accenting the vowel creates an ambiguity: either the vowel is in narrow focus, or the entire syllable is in broader focus through integration.

We approached the problem by examining the production and perception of identical (monosyllabic) words containing narrow focus contrasts, involving individual segments, and broader focus involving the entire word, as in (3a–d):

(3) a. I said pit, not bit [contrasted element: C1]
b. I said pit, not pat [contrasted element: V]
c. I said pit, not pick [contrasted element: C2]
d. I said pit, not back [contrasted element: entire syllable]

If the syllable is truly the smallest prosodic domain, the phonetic structure of the word *pit* in (3a–d) should be the same, irrespective of the position or scope of the contrasted unit. However, if there are systematic differences in the four realizations of *pit* due to variation in focus (here: contrast), the segment rather than the syllable is the minimal prosodic domain. Moreover, if (3a–c) differ from each other, but (3b) does not differ from (3d), the ambiguity between (3b) and (3d) is evidence that the vowel is the prosodic head of the syllable.

If segmental contrasts are prosodically coded, and if the vowel can be shown to be the exponent of the syllable, the most elegant account of accent is that it is basically a property of the vowel, with the option of marking ever larger domains for focus (syllable, word, constituent, phrase, etc.) through the mechanism of integrative focus.

The perception experiment and subsequent stimulus analysis to be described in the following sections were designed to explore these possibilities.

6.2 Perception of subsyllable contrasts

6.2.1 Introduction

Speakers were asked to read out materials with target words of the CVC type placed in contexts that suggested a narrow focus contrast on either C1, V, or C2, or with the entire syllable contrasted (relatively broad focus). The primary purpose of the experiment was to establish to what extent listeners would be able to retrieve the focus distribution intended by the speaker, when the target words were presented after having been isolated from their original, spoken context. If listeners could correctly decode the intended focus distribution from the spoken stimulus, or at least perform this task well above chance, this would mean that prosody is used to focus linguistic units below the level of the syllable, i.e. individual segments.

Secondly, we were interested in testing the consequences of the status of the vowel as the exponent (Fuchs 1984), or prosodic head, of the syllable. As explained in section 6.1.1, marking the head of a prosodic unit for focus is always ambiguous for the listener, since it can be construed either as narrow focus on the exponent itself, or as broader focus on the larger constituent which is headed by the exponent through so called integrative focus. If listeners can differentiate between intended narrow focus on C1, V and C2, but not between narrow focus on V and broader focus on the entire target syllable, this would be independent support for the claim that the vowel is the prosodic head of the syllable.

6.2.2 Method

Three Dutch target words were selected, each a CVC monosyllable with a phonologically long vowel and voiced consonants throughout: *boon* [bo:n] "bean," *vaar* [va:r] "sail" and *zeem* [ze:m] "sponge." Each target was embedded in a fixed carrier sentence in prefinal position, in which it was

contrasted with an earlier word, e.g.:

(4) Ik heb niet b*ee*n maar b*oo*n gezegd
ɪk hɛp ni:t be:n ma:r bo:n ɣəzɛχt]
(I have not leg but bean said)

In the above example the contrasted element is the vowel. Likewise, the contrasted elements could be C1, C2, or the entire syllable.

(5) Ik heb niet *w*oon maar *b*oon gezegd [contrasted element C1]
Ik heb niet boo*m* maar boo*n* gezegd [contrasted element C2]
Ik heb niet *veer* maar *boon* gezegd [entire syllable contrasted]

Five native Dutch speakers (two males, three females), fully naive to the purpose of the experiment, read out the twelve utterances. Speakers were seated in a sound-insulated recording booth, and were recorded on audio tape using semi-professional equipment (Sennheiser MKH-416 condenser microphone, Studer-Revox B77 recorder).

The recordings were analog-to-digital (A/D) converted (10 kHz, 12 bits, 0.3–4.5 kHz BP) and stored on computer disk. This filter band was chosen to prevent aliasing and to de-emphasize the strong energy concentration in the region of the fundamental. The last three words of each sentence (. . . *maar boon/zeem/vaar gezegd*) were excised from their spoken context using a high resolution waveform editor. These fragments were D/A converted and recorded back on to audiotape in quasi-random order (excluding immediate succession of the same lexical target word). Each stimulus was recorded three times in a row with three-second intervals (offset to onset). Triplets were separated by a seven second interval. The stimulus set proper was preceded by five practice triplets using similar, but not identical, CVC words.

The tape was presented over loudspeakers in a quiet lecture room to nine native Dutch listeners (staff and/or students at the Department of Linguistics/Phonetics Laboratory of Leiden University). Listeners indicated what they thought would be the most likely context for each stimulus, with forced choice from among four alternatives. The answer sheets listed the stimulus words in the order in which they appeared on the tape. Each target word was printed in four alternative sentences with C1, V, C2, or the entire target syllable in contrast, as exemplified above.

6.2.3 Results

The dataset nominally comprised 540 responses (9 listeners * 5 speakers * 12 stimulus types). In nine cases (1.7%) listeners failed to respond, so that the actual number of responses was 531. The data analysis will proceed in two

stages. We shall first examine the results for all responses. This analysis will reveal a number of tendencies, not all of which can be shown to be statistically significant. In a further analysis, however, we shall select the more task-proficient speakers and listeners. After this selection, the trends that are visible in the aggregate data can easily be shown to reach statistical significance.

6.2.3.1 First analysis: all data

Table 6.2.1 presents a confusion matrix with the four intended focus distributions vertically and the listeners' reconstruction of the speakers' intention horizontally. Generally, the effects of intended focus distribution are small. Nevertheless the response distributions deviate highly significantly from chance for each of the four stimulus conditions by a simple chi-square test. Moreover, the overall number of correct responses (the main diagonal cell frequencies taken together) is significantly better than chance ($p = 0.001$, binomial test). When the speaker intended to contrast C1, C1 is the most frequent response category (37%). When V is the contrasted element, V is the most frequently chosen option (38%). When C2 is focused, it is the second most frequent option in its row (30%). However, there appears a rather strong bias throughout the matrix against C2 (and favoring V); when considered column-wise, the C2 cell on the main diagonal contains about as many responses as the rest of the column taken together. Finally,

Table 6.2.1. *Confusion matrix of focus distribution as intended by speakers (C1 contrasted, V contrasted, C2 contrasted, whole word contrasted) and as perceived by listeners. Absolute numbers and row percentages are indicated.*

Intended contrast	Perceived focus distribution			
	C1	V	C2	Word
C1	49	44	19	21
	37%	33%	14%	16%
V	39	51	11	32
	29%	38%	8%	24%
C2	22	44	40	26
	17%	33%	30%	20%
Word	30	56	14	33
	23%	42%	11%	25%

Table 6.2.2. *Pairwise chi-square comparisons of focus conditions (rows in table 6.2.1). Df = 3 for all comparisons.*

Focus conditions compared	Chi-square	p-value
C1 vs. V	6.1	0.108
C1 vs. C2	18.3	< 0.001
C1 vs. Word	9.4	0.024
V vs. C2	22.4	< 0.001
V vs. Word	1.8	0.619
C2 vs. Word	16.0	0.001

when the intended contrast is on the whole word, focus is predominantly perceived on V (42%), with the whole-word option in second place (25%).

The data show that the response distributions generally differ significantly for all intended focus conditions, except for the pair "V in focus" versus "whole word in focus." This can be observed in table 6.2.2, which lists the results of six pairwise comparisons between rows in the confusion matrix, using chi-square tests (df = 3). Although the first comparison (C1 versus V) does not yield a truly significant difference (but see below), this difference is a trend at least. However, the difference between V and whole word in focus is absolutely insignificant. In sum, the results indicate a weak, but significant, effect of intended focus distribution. Listeners are able to some extent, and above chance level, to reconstruct the speaker's intention from the acoustic make-up of the stimulus. The response distributions for "focus on C1" and "focus on C2" differ considerably, and in the predicted direction. Both these types differ from either "focus on V" or from "focus on whole word," but these latter two do not differ from each other.

6.2.3.2 Second analysis: selection of data

Speakers and listeners may well differ in their abilities to encode and decode subtle differences in focus distribution. Let us therefore examine the individual performance of speakers and listeners. For the sake of conciseness, we shall present only the percentage of correctly transmitted focus distributions, broken down by speaker (table 6.2.3a) and by listener (table 6.2.3b), rather than presenting complete confusion matrices. We observe that four listeners (#1, #4, #6 and #9) performed their task better than the others. Only two speakers (#4 and #5) were able to more or less successfully encode differences in focus distribution. Let us go through the data twice more, once after selecting only the four most able listeners (but across all speakers), and once after selecting only the two best speakers (but

Table 6.2.3. *Percentage of correct responses broken down by individual listeners (panel a) and by individual speakers (panel b)*

(a)	Listener #	Mean	Cases
	1	0.3667	60
	2	0.2667	60
	3	0.3462	52
	4	0.3667	60
	5	0.2833	60
	6	0.3667	60
	7	0.3220	59
	8	0.2500	60
	9	0.3667	60
(b)	Speaker #	Mean	Cases
	1	0.2547	106
	2	0.2308	104
	3	0.2500	108
	4	0.5185	108
	5	0.3714	105
For entire population		0.3258	531

Table 6.2.4. *Like table 6.2.1; only the results for the four best listeners have been included.*

Intended contrast	Perceived focus distribution			
	C1	V	C2	Word
C1	23	17	10	10
	38%	28%	17%	17%
V	15	22	4	19
	25%	37%	7%	32%
C2	7	15	21	17
	12%	25%	35%	28%
Word	14	15	9	22
	23%	25%	15%	37%

Table 6.2.5. *Pairwise chi-square comparisons of focus conditions (rows in table 6.2.4). Df = 3 for all comparisons.*

Focus conditions compared	chi-square	p-value
C1 vs. V	7.7	0.053
C1 vs. C2	14.4	0.002
C1 vs. Word	6.9	0.076
V vs. C2	15.9	0.001
V vs. Word	3.5	0.321
C2 vs. Word	7.8	0.051

across all listeners). The confusion matrices and associated statistics resulting after this selection are presented in tables 6.2.4 and 6.2.5 (listener selection), and tables 6.2.6 and 6.2.7 (speaker selection). Concentrating on percent correct (main diagonal cells) we observe – predictably – that the intended focus distribution has been transmitted more effectively than in table 6.2.1. Also, the selected listeners suffer less from bias. The intended contrast has been retrieved at 12% above chance level, with clear differences between the response distributions for C1, V, and C2, but, again, with no

Table 6.2.6. *Like table 6.2.1; only the results for the two most successful speakers have been included.*

Intended contrast	Percieved focus distribution			
	C1	V	C2	Word
C1	30	13	7	4
	56%	24%	13%	7%
V	17	23	3	11
	32%	43%	6%	20%
C2	5	17	23	7
	10%	33%	44%	14%
Word	11	21	2	19
	21%	40%	4%	36%

Table 6.2.7. *Pairwise chi-square comparisons of focus conditions (rows in table 6.2.6). Df = 3 for all comparisons.*

Focus conditions compared	chi-square	*p-value*
C1 vs. V	11.2	0.011
C1 vs. C2	27.7	< 0.001
1 vs. Word	23.2	< 0.001
V vs. C2	23.7	< 0.001
V vs. Word	3.7	0.296
C2 vs. Word	25.8	< 0.001

clear distinction between "focus on V" and "focus on the whole word." Pairwise comparisons of focus conditions bear this out (table 6.2.5). More pairwise contrasts reach statistical significance after than before listener selection, including the contrast C1 versus V that could not be shown to be significant in table 6.2.2. Unfortunately, the contrast "focus on C1 versus Word" that was significant in table 6.2.1 now just falls short of reaching significance. The crucial point, of course, is that only one contrast remains absolutely insignificant, viz. "focus on V versus Word."

The confusion matrix for the two best speakers is presented in table 6.2.6 with pairwise contrasts in table 6.2.7. After the two most task-proficient speakers have been selected, all pairwise comparisons yield significant differences, with one exception: "focus on V versus whole word," which difference remains totally insignificant.

6.2.4 Conclusion

Clearly then, some speakers are much more proficient in encoding differences in focus distribution at the level of the segment than others. Also, certain listeners are better attuned to these cues than others. However, especially when the better performers have been picked out, the results show that contrasts (narrow focusing) on linguistic units below the level of syllable, i.e. individual segments, can be made with some measure of success, and certainly above chance. Moreover, given that differences could nowhere be established between "focus on V" versus "focus on whole word," these data independently support the status of the vowel as the exponent or prosodic head of the syllable.

6.3 Acoustic analysis of subsyllable contrasts

6.3.1 Introduction

What cue or cues do speakers use to convey subsyllable contrasts? To answer this question we acoustically analyzed the speech material used in the perception experiment. We had originally assumed that our speakers would use rather trivial tricks to express narrow focus on individual segments, such as making the contrasted segment unnaturally long or loud. One part of the stimulus analysis is therefore concentrated on measures of absolute and relative segment duration and intensity. However, we have also looked for more subtle, and less trivial, cues in the position and shape of the accent-lending pitch movements on the syllables that contained the various contrasts. In the next section (6.3.2) we shall outline the types of acoustic analysis that were performed; the results and preliminary conclusions will be presented in sections 6.3.3 and 6.3.4, respectively.

6.3.2 Analysis

After A/D-conversion (see p. 81 above) the target phrases were submitted to a pitch-extraction and tracking algorithm using the method of subharmonic summation (Hermes 1988) which calculated F_0 for time frames of 10 ms. Remaining errors (typically octave jumps) were corrected by hand.

For each target phrase acoustic properties were measured in six domains: (i) segment duration, (ii) duration of pitch movements, (iii) excursion size of pitch movements, (iv) synchronization of pitch movements relative to segment boundaries, (v) segment intensity, and (vi) spectral distribution. In the next sections we shall discuss the various measurements per domain.

6.3.2.1 Segment durations

Segment durations were measured by hand (eye) using a high-resolution waveform editor. Segment boundaries were determined by visual criteria only, i.e. abrupt changes in the amplitude and shape of successive glottal periods. In order to define valid relative duration measures some additional time intervals were determined, yielding the following set of duration measurements:

duration of entire sentence
duration of sentence until target CVC word
duration of target segments:
  initial consonant C1
  medial vowel V
  final consonant C2

6.3.2.2 Duration of pitch movements
We started from the assumption that each contrastive accent would be realized as a so-called pointed-hat pitch configuration (configuration 1&A in the intonation grammar of Dutch, cf.'t Hart *et al.* 1990). However, we anticipated that the rise and the fall constituting this configuration could be separated by a plateau. The pitch contour of each target was therefore reduced to three straight lines (in a log frequency by linear time display), whose durations were measured with a precision of 10 ms:

duration of pitch rise
duration of pitch plateau
duration of pitch fall.

6.3.2.3 Excursion size of pitch movements
The F_0 intervals between the onset and offset moments of pitch rises and falls were measured in semitones. Semitone conversion abstracts away from actual pitch levels, enabling better comparison between speakers. The following excursion sizes were determined:

excursion size of pitch rise
excursion size of pitch fall.

6.3.2.4 Synchronization of pitch movements relative to segment boundaries
We determined the moments of onset and offset of the accent-lending rise and fall for each target word, expressed in milliseconds relative to the vowel onset. When a pitch-movement onset or offset was located before the vowel onset, a negative value resulted. The following set of synchronization measures was determined:

onset of (virtual) pitch rise
offset of pitch rise
onset of fall
offset of fall.

6.3.2.5 Segment intensity
The peak intensity of each target segment was measured in decibels (25.6 ms integration time). Since there is no guarantee that our speakers observed a constant distance to the microphone, the segment intensities were expressed as differences (in dB) relative to a reference vowel that occurred outside the target word in the spoken context: [ɛ] in the final word *gezegd* [ɣəzɛχt] (this is the last nonreduced vowel within the same phonological phrase as the target). If the reference has a lower intensity than the target segment, the

difference was given a negative value. The following intensities were entered in the database:

relative peak intensity of initial consonant C1
relative peak intensity of medial vowel V
relative peak intensity of final consonant C2.

The intensity of the initial voiced stop [b] of the target word *boon* could not be measured, so the number of observations for this parameter is limited to 10 rather than 15.

6.3.2.6 Spectral distribution

At the intensity maximum of each target vowel the center frequencies and bandwidths of the lowest five formants (F_1 through F_5, B_1 through B_5) were estimated by the split-Levinson LPC-based robust formant tracking method (Willems 1986; analysis window 25.6 ms, time-shift 10.0 ms). Only F_1 and F_2 were used for further analysis.

6.3.3 Results

6.3.3.1 Segment durations

It would seem a reasonable assumption that narrow focus on one segment would prompt the speaker to lengthen this segment relative to its competitors in the same syllable.

Table 6.3.1 presents the absolute segment durations of the initial consonant (C1), the vowel (V), the final consonant (C2), and the duration of the entire target word (W), broken down by intended contrast condition

Table 6.3.1. *Absolute (in ms) and relative (in percent) duration of initial consonant (C1), medial vowel (V), final consonant (C2), and of entire word, broken down by focus condition: "focus on initial consonant" (C1), "focus on medial vowel" (V), "focus on final consonant" (C2), and "focus on entire word" (Word). Data have been accumulated over lexical items and speakers; each mean is based on 15 measurements.*

Focus on	Absolute duration of				Relative duration of			
	C1	V	C2	Word	C1	V	C2	Word
C1	100	176	100	375	26	47	27	16
V	97	166	106	369	26	45	29	16
C2	95	180	106	382	25	48	27	16
Word	100	166	99	366	27	46	27	16

("C1 in focus," "V in focus," "C2 in focus," "whole word in focus"). The data have been accumulated over speakers and over lexical items. Table 6.3.1 further contains relative segment durations that express the duration of individual segments as a percentage of the duration of the entire word, and the duration of the word as a percentage relative to the duration of the entire utterance. None of the segment durations, whether absolute or relative, is influenced by a difference in focus condition. Classical two-way analyses of variation (ANOVAS), performed separately for each of the acoustic measures with focus condition and speaker as fixed factors, show that all effects of focus distribution are completely insignificant, $F_{(3,56)} < 1$.

6.3.3.2 Duration of pitch movements

The duration of the three components of the accent-lending pitch movement (rise, high plateau, fall) is presented in table 6.3.2, accumulated over speakers and lexical items, but broken down by focus condition.

We reasoned that narrow focus contrasts might be pointed out to the listener when the speaker makes the rise–fall combination more compact in time, i.e. with shorter (and steeper) rises and falls, and centered over the contrasted segment. So we expected that pitch configurations would be shifted along the time axis depending on the position of the contrasted segment (see below under synchronization measures), and that a narrow focus contrast would be characterized by a more compact shape of the pitch configuration. Whether this is true can be examined in table 6.3.2. The pitch rise lasts longer as the focused segment is closer to the left word edge, i.e. the pitch rise is long when C1 is in focus, average when either V or the whole word is in focus, and shortest when C2 is focused, $F_{(3,56)} = 4.2$, $p = 0.011$.

The duration of the high plateau is quite short throughout, and differences between the focus conditions cannot possibly reach perceptual

Table 6.3.2. *Duration (in ms) of pitch rise, high plateau, and pitch fall on target word, broken down by focus condition (as in table 6.3.1).*

Focus on	Duration of pitch movement		
	rise	plateau	fall
C1	254	21	166
V	213	31	160
C2	187	25	183
Word	203	43	145

relevance, even though ANOVA shows significant effects for this parameter, $F_{(3,56)} = 3.1$, $p = 0.039$.

The duration of the pitch fall is about 20 ms longer than average when C2 is in focus, and some 20 ms shorter than average when the whole word is in focus. This effect, however, fails to reach statistical significance, $F_{3,56} = 1.5$, n.s.

Thus it would appear that the pointed-hat configuration is more distributed in time when the subsyllable contrast is towards the end of the target syllable, and more compact when the contrast is towards the beginning of the syllable.

6.3.3.3 Excursion size

We expected that accenting a constituent that does not normally receive accent, i.e. that is not the exponent of its larger domain, would prompt speakers to give extra prominence to this accent. For instance, it would seem plausible that accents on syllables that are not lexically stressed (as in the phrase *putting the emPHAsis on the wrong sylLABle*) are given extra prominence by increasing the magnitude of the pitch excursions. If this reasoning is correct, we should observe larger pitch excursions in our material when the narrow-focus contrast does not involve the exponent of the syllable, i.e. involves the consonant segments (C1 or C2), than when it is on the exponent, i.e. on the vowel or on the whole word.

Table 6.3.3 presents mean excursion size of pitch rise and fall on the target word, across speakers and lexical items, but broken down by focus condition. Here we shall consider the results for all speakers; the results for

Table 6.3.3. *Excursion size of pitch rise and of pitch fall (in semitones) on target word, broken down by focus condition (as in table 6.3.1). In the rightmost two columns, the breakdown is repeated for the two most successful speakers (means are now based on 6 measurements).*

Focus on	Excursion size of pitch movements			
	all speakers		two best speakers	
	rise	fall	rise	fall
C1	6.6	9.0	9.0	10.9
V	5.3	8.0	7.0	8.5
C2	6.4	9.4	7.1	10.0
Word	6.0	7.6	8.2	8.3

the two most successful speakers have been included separately for the sake of the general discussion only (section 6.4.1). The excursion size of the rise is larger when there is narrow focus on either C1 or C2 than when focus is on either the vowel (1 semitone difference) or the whole word (0.5 semitone difference). Unfortunately, this effect of focus on excursion size of the rise just falls short of statistical significance, $F_{(3,56)} = 2.6$, $p = 0.067$. There is a similar effect of focus condition on the excursion size of the fall: 1 to 2 semitone larger falls are observed for focus on consonants than for focus on vowel or whole word. Here the effect is just significant, $F_{(3,56)} = 2.8$, $p = 0.051$ (with lexical word as a second factor and after removing differences between speakers through normalization by Z-transformation).

Still, to us these findings suggest that focusing on the consonants of the syllable is "marked" by a more conspicuous pitch movement than in the normal situation, when focus is on the head of the syllable, i.e. the vowel.

6.3.3.4 Synchronization of pitch movements relative to segments

We expected our speakers to center the rise–fall configuration over the specific segment they wished to put in focus position. Accordingly we predicted that the pivot points of the pitch contour, especially the middle two (end of rise, beginning of fall) that are associated with the pitch peak, would shift along the time axis with the position of the focused segment.

Table 6.3.4 presents the relative positions of onset and offset of the accent-lending pitch rise and fall, expressed in ms relative to the vowel onset of the target word.

We observe a tendency especially for the middle two pivot points in the F_0 contour (i.e. the culmen or pitch peak) to be shifted along the time axis into the direction that is opposite to the position of the focused segment within

Table 6.3.4. *Synchronization of pitch movements (rise onset, rise offset, fall onset, fall offset), in ms relative to vowel onset broken down by focus condition (as in table 6.3.1).*

Focus on	Synchronization point of pitch movement			
	rise onset	rise offset	fall onset	fall offset
C1	−163	91	112	278
V	−163	51	82	241
C2	−147	39	64	247
Word	−145	58	101	246

the syllable. When C1 is in focus, the pitch peak is shifted towards the end of the syllable; when the final consonant C2 is in focus, the culmen is advanced towards the beginning of the target syllable. The pitch contour assumes a middle position when either V or the entire word is in focus. The effect is most regular for the fall onset, i.e. the position along the time axis where the culmen of the pitch contour is located. For this parameter the effect of focus position reaches significance by a classical two-way ANOVA with focus condition and lexical word as fixed factors, but only when the two best speakers are selected, and after speaker normalization through Z-transformation, $F_{(3,20)} = 3.8$, $p = 0.054$.

6.3.3.5 Intensity

An easy way for the speaker to mark an individual segment for contrast would be to increase its intensity. Table 6.3.5 contains the peak intensities of C1, V, and C2 expressed in decibels above the peak intensity of the reference vowel (see p. 88 above). As before, data have been accumulated across speakers and lexical items, but are broken down by focus condition. When the initial consonant is in focus, its relative intensity is slightly stronger than in other focus conditions. Similarly, when the vowel is in focus, it is somewhat more intense than when it is not. However, the effects are minute, and this tendency is reversed in the case of focus on the final segment, so that there is no general effect of focus position on the intensity of individual segments, $F_{(3,56)} < 1$ for all intensity parameters.

Table 6.3.5. *Intensity of initial consonant (C1), medial vowel (V), and final consonant (C2), expressed in decibels relative to the intensity of the last vowel in the utterance, broken down by focus condition (as in table 6.3.1). Means for the two rightmost columns are based on 15 measurements, the mean for the leftmost column is based on 10 measurements.*

Focus on	Relative intensity of		
	C1	V	C2
C1	6.2	11.6	2.8
V	5.1	12.1	2.9
C2	3.7	10.1	0.9
Word	4.5	11.8	2.2

Table 6.3.6. *Center frequency of first and second formants (in Hz) broken down by word and by focus condition (as in table 6.3.1). Each mean is based on five measurements.*

Focus on	First formant of:			Second formant of:		
	/bo:n	va:r	ze:m/	/bo:n	va:r	ze:m/
C1	480	741	447	957	1235	1719
V	495	682	423	974	1202	1549
C2	497	688	449	985	1198	1486
Word	501	731	416	975	1224	1554

6.3.3.6 Spectral distribution

Unaccented segments are generally articulated less carefully, which leads to temporal (see p. 89 above) and spectral reduction, so that peripheral vowels tend to gravitate towards the center of the F_1/F_2 plane. Table 6.3.6 therefore presents the center frequencies of F_1 and F_2 broken down by word and by focus condition. Separate two-way analyses of variance on all formants (only F_1 and F_2 are shown) with vowel type (/e:, o:, a:/) and focus condition (C1, V, C2, Word) showed complete insignificance of focus shifts, $F_{(3,56)} < 1$ for all parameters, as well as utter insignificance of any vowel by focus interaction, $F_{(6,48)} < 1$ for all formants. Clearly, spectral differences do not cue our subsyllable contrasts.

6.3.4 Conclusions

Virtually none of the large number of acoustic parameters measured or derived proved susceptible to effects of narrowing and/or shifting focus on individual segments within a syllable. Focusing on individual segments has no effect on either the duration, intensity, or spectral characteristics of segments, even though these would be the most likely candidates for focus cues on the segmental level.

However, there are systematic effects of subsyllable focus shifts on the position and shape of the accent-lending pitch contour on the target word. Typically the position of the pitch peak moves away from the center of the syllable in such a way that it assumes a position that is opposite that of the focused segment, i.e. late when C1 is in focus, intermediate when V or the whole word is focused and early when C2 is in focus. We shall come back to this in the general discussion (section 6.4). Also, it seems that the rise is shorter (and steeper) when the pitch peak occurs early in the syllable (C2 in

focus), and longer and more gradual when the peak occurs late (C1 in focus). Finally, the rise has a larger excursion when a consonant is in focus than when the vowel (or the whole word) is in focus.

6.4 General discussion

6.4.1 Summary of main findings

The purpose of this study was to find experimental support for the hypothesis that the segment rather than the syllable is the smallest, and basic, domain of a (pitch) accent. We approached this issue by examining the production and perception of relatively unusual speech utterances containing contrastive elements at the level of individual segments. Although at first sight this may seem a highly contrived communicative situation, I must stress that there is no other way if we want to get at the true nature of accent. The fact that the crucial events only occur in exceptional communicative situations may well be the reason why no one has pursued the possibility of accent as a segmental property before.

Our experiment demonstrates that at least some speakers have the means to express narrow focus on linguistic units below the level of the syllable. Crucially, such speakers do this by purely prosodic means, viz. by changing properties of the accent-lending pitch contour (its shape and location) on the syllable that contains the contrasted segment, rather than by changing acoustic properties of the individual contrasted segment. Moreover, both the results of the perception experiment and of the stimulus analysis clearly bear out that narrow focus on the vowel is brought about by the same means as broader focus on the entire syllable, which supports the status of the vowel as the prosodic head of the syllable.

The effects of focus distribution within the syllable are subtle. It takes a highly proficient speaker to produce them, but if he does, at least certain listeners are able to reconstruct the speaker's intended focus distribution much better than by chance. It appears that the best speakers exploit intonational means more fully than ordinary speakers do. As a case in point, table 6.3.3 shows that the optimal speakers used larger F_0 excursions to mark accents than ordinary speakers. Note, incidentally, that neither speakers nor listeners were trained, or given much time to develop an ad-hoc strategy for marking subsyllable contrasts. We are therefore convinced that we have studied phenomena with linguistic significance, rather than experimental artifact.

Our general conclusion is therefore that accent is best regarded as a property of the segment. Stress, of course, will remain what it has always been: an

abstract property of a word specifying the syllable that has the integrative accent position within the word domain.

The true nature of accent will only come to light in exceptional communicative situations, such as those used in the present experiment, in which a speaker wishes to focus on one specific consonant within a syllable. Normally, however, the accent will be on the head of the syllable, i.e. the vocalic nucleus, so that the entire syllable will be highlighted through integrative focus. Since, again, it takes unusual circumstances for the accent not to occur on the lexically stressed syllable, an accent on the vowel generally marks the whole (polysyllabic) word for focus, and so on for larger domains above the word level.

6.4.2 Pitch-peak location and perceived duration

We were both amazed and puzzled by the finding that the pitch peak of the accent marking a segmental contrast should tend to move away from the middle of the syllable in a direction opposite to the location of the contrasted segment, rather than coincide with it. On second thoughts, however, this behavior may not be so odd as it seems. Normally, the pitch peak coincides with the vocalic nucleus, i.e. is located roughly halfway through the syllable. By postponing the pitch peak (either by moving the entire rise–fall configuration towards the end of the syllable or by making the rise longer as well) the speaker creates the impression that the segment(s) preceding the pitch peak last longer, and those following it are shorter. When the pitch peak is advanced towards the syllable onset the listener is tricked into believing that the first half of the syllable is short and the second half long. If this hypothesis is correct, shifting the pitch peak is used by speakers as an alternative to manipulating segment durations within the syllable.

There is circumstantial evidence for the correctness of this account. Van Dommelen (1980) made a contrastive study of production and perception of vowel duration in Dutch and German. The impressionistic and pedagogic literature claims that German high vowels are longer than their Dutch counterparts, and suggests that Dutch learners of German should be taught to double the duration of these vowels. Van Dommelen's measurements, however, brought to light the fact that German high vowels were not longer than their Dutch counterparts, and that the difference in perceived duration could not be explained by production duration. Adriaens (1991) showed that there are systematic intonation differences between Dutch and German, not only in the excursion size of the pitch movements, but also in the timing of the accent-lending rise. Crucially, the German accent-lending rise starts very late in the syllable, whereas the

Dutch rises occur early. Now, if it is true that a late accent-lending rise makes the preceding part of the syllable sound long, Van Dommelen's paradox is solved.

6.4.3 Final remarks

Our experiment is a small-scale exploration that leaves room for improvements and extensions. Its findings will have to be replicated with larger groups of speakers and listeners. The stimulus analysis will have to be submitted to more sophisticated statistical procedures. So far only gross relations have been established between perception of intended focus distribution and acoustic differences. Rather we should try to correlate the production and perception of subsyllable focus differences on a token-individual basis. And ultimately, we shall have to check whether the acoustic differences in shape and position of the pitch configuration, rather than other differences, are indeed the perceptual cues that listeners use to reconstruct the intended subsyllable focus distribution. This will necessarily involve systematic manipulation of selected acoustic parameters through speech synthesis or resynthesis techniques, and testing the perceptual effects of such manipulations.

Acknowledgment

The experiment and stimulus analysis described in this paper were run by my student Jacqueline de Leeuwe as part of a class assignment. I thank my colleagues Anneke Neijt (Department of Dutch, Leiden University), Tom Cook (Department of African Linguistics, Leiden University), and Reneé van Bezooijen (Department of Linguistics, Nijmegen University) for ideas and discussion.

References

Adriaens, L. 1991. Ein Modell deutscher Intonation, eine experimentell-phonetische Untersuchung nach den perzeptiv relevanten Grundfrequenzänderungen in vorgelesenem Text. Ph.D. dissertation, Eindhoven University of Technology.

Baart, J.L.G. 1987. Focus, syntax and accent placement. Ph.D. dissertation, Leiden University.

Bolinger, D.L. 1958. A theory of pitch accent in English. *Word* 14: 109–149.

1961. Contrastive accent and contrastive stress. *Language* 37: 83–96.

Chomsky, N. and M. Halle. 1968. *The Sound Pattern of English.* New York: Harper & Row.

Dommelen, W.A. van. 1980. Temporale Faktoren bei ausländischem Akzent, eine kontrastive deutsch-niederländische Untersuchung zur Produktion und Perzeption von Segmentdauerwerten. Ph.D. dissertation, Leiden University.

Fuchs, A. 1984. "Deaccenting" and "default accent". In Gibbon, D. and H. Richter (eds.) *Intonation, Accent and Rhythm.* Berlin: de Gruyter 134–164.

Gussenhoven, C. 1984. *On the Grammar and Semantics of Sentence Accents.* Dordrecht: Foris.

Hart, J.'t, R. Collier and A. Cohen. 1990. *A Perceptual Study of Intonation.* Cambridge: University Press.

Hermes, D.J. 1988. Measurement of pitch by subharmonic summation. *Journal of the Acoustical Society of America* 83: 257–264.

Ladd, D.R. 1980. *The Structure of Intonational Meaning: Evidence from English.* Bloomington: Indiana University Press.

Nooteboom, S.G. and J.G. Kruyt. 1987. Accents, focus distribution, and the perceived distribution of given and new information. *Journal of the Acoustical Society of America* 82: 1512–1524.

Sluijter, A. 1992. Lexical stress and focus distribution as determinants of temporal structure. In R. Bok-Bennema and R. van Hout (eds.) *Linguistics in the Netherlands.* Amsterdam: John Benjamins, 247–259.

Willems, L.F. 1986. Robust formant analysis. *IPO Annual Progress Report* 21. Eindhoven: Institute for Perception Research, 34–40.