Prominence of pitch-accented syllables

PROMINENCE OF PITCH-ACCENTED SYLLABLES

PERCEPTIEVE OPVALLENDHEID VAN LETTERGREPEN MET EEN ACCENTVERLENENDE TOONHOOGTEBEWEGING

PROEFSCHRIFT

ter verkrijging van graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de Rector Magnificus, prof.dr. M. Rem, voor een commissie aangewezen door het College van Dekanen in het openbaar te verdedigen op maandag 18 november 1996 om 16.00 uur

door

Hans Hardy Rump

geboren te Bremen

Dit proefschrift is goedgekeurd door de promotoren:

prof.dr. R. Collier
en
prof.dr. S. Nooteboom (Univ. Utrecht)

Copromotor: dr. D. Hermes

Printed by: PrintPartners Ipskamp (voorheen CopyPrint 2000), Enschede

Gegevens:

Rump, Hans Hardy

Prominence of pitch-accented syllables = Perceptieve opvallendheid van lettergrepen met een accentverlenende toonhoogtebeweging / Hans Hardy Rump. - Eindhoven : Technische Universiteit Eindhoven, 1996. - Fig., tab.
Met lit.opg.
ISBN 90-386-0089-5
Trefw.: perceptie / prosodie (taalkunde) / accentuering

Voor Hannie
Maaike en Elise

Paranimfen:

Cynthia Grover
Margreet Sanders

CONTENTS

1 GENERAL INTRODUCTION

1.1 PITCH ACCENTUATION AND PROMINENCE

The prominence of pitch-accented syllables is an important factor in speech perception. By highlighting the most informative parts of an utterance, prominence helps listeners in determining its meaning. In addition, prominence influences well-formedness or naturalness of an utterance: a wrong degree of prominence can be very conspicuous and degrades speech quality. It is well known that the pitch contour realized on an utterance determines to a large extent the prominence of the syllables which constitute the utterance. In this thesis we have studied how the prominence of pitch-accented syllables is related to the pitch contour of an utterance.

First, some of the terminology used in this thesis needs to be explained. The most important terms are: 'pitch', 'pitch accent', and 'prominence'.

Pitch is tonal sensation which, for speech, is mainly related to (quasi-)periodicity of the signal. The inverse of this period is most often referred to as fundamental frequency (F_0). 'Pitch manipulations' are manipulations of the periodicity that have direct audible effects on the pitch of the signal. Pitch can be measured on various scales. Here, the three most relevant scales are the Hertz scale, the semitone scale, and the ERB-rate scale. The Hertz (Hz) scale is a linear scale, the semitone (st) scale is a logarithmic scale. Musical intervals like octaves are equal when expressed in st, but different when expressed in Hz. The ERB-rate (Equivalent-Rectangular-Bandwidth-rate) scale is psychophysically determined and based on measurements of auditory bandwidths (see Patterson, 1976; Glasberg and Moore, 1990). Between 50 and 500 Hz, the pitch range of normal speech, the ERB-rate (E) scale is between the linear Hz scale and the logarithmic st scale. The graphical representation of the pitch values of an utterance is called "pitch contour" (for an example see fig. 1).

Two terms which are often used interchangeably are the terms **stress** and **(pitch) accent**. In our view, accent is a phonetic feature at utterance level and is primarily related to pitch. Stress is a phonological feature at word level and marks the syllable which is most likely to be (pitch-)accented if a word is spoken in citation form (i.e. in isolation, e.g.: ˌisoˈlation, ˈ indicating main stress and ˌ indicating secondary stress). In an utterance only a few (main) stressed syllables are accented, depending on linguistic factors such as discourse structure.

The **prominence** of a syllable is defined as its perceptual conspicuousness or salience relative to the neighbouring syllables. Several factors are known to contribute to the prominence of pitch-accented syllables: presence of a pitch movement, together with other, stress-related factors: duration, amplitude, and vowel quality. Fry (1955, 1958) has already shown that the contribution of the last three variables to prominence is far smaller than that of the presence of a pitch movement. The main goal of this thesis is to clarify the relationship between pitch and prominence of pitch-accented syllables.

Other terms which will be frequently used are 'declination', 'baseline', 'topline', and 'excursion size'. Since these terms are specific for the method of intonation research at the Institute for Perception Research, this method will be introduced briefly (for an extended overview see 't Hart, Collier, and Cohen, 1990). One of its most important features is the method of "close-copy stylization": an original pitch contour is replaced by a pitch contour consisting of as few straight lines as possible without introducing audible differences between the original pitch contour and the stylized. The stylized one is then perceptually equal to the original contour (for a detailed description of the stylization method, see De Pijper, 1983, for a scheme see 't Hart *et al.*, 1990, p. 65). By stylizing large quantities of speech materials, regularities in the course of intonation were found. These were then formalized into intonation rules. After synthesizing artificial pitch contours, it was perceptually tested whether these contours were perceptually equivalent to the original ones, i.e. whether they conveyed the same meaning. One of the main results was the finding that the large variety of pitch movements found in natural speech can be represented by a rather limited set of so-called 'standard pitch movements' consisting of accent-lending and non-accent-lending pitch movements connected by stable or slightly sloping pitch. The physical differences between accent-lending and non-accent-lending pitch movements are their timing within a syllable and the duration of a movement itself: accent-lending pitch movements are relatively short (standard duration: 120 ms) and are generally located in the vowel nucleus of the accented syllable. Summarizing, a stylized pitch contour of an utterance consists of the perceptually relevant pitch movements together with parts in which the pitch stays relatively stable (see 't Hart *et al.*, 1990). This is illustrated in Fig. 1.

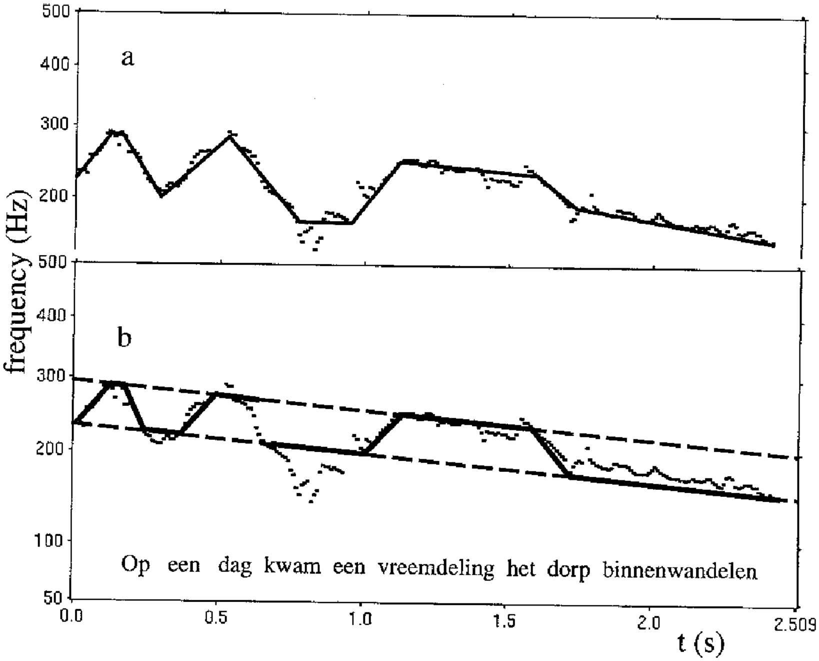


Figure 1. *Original (dotted line) pitch contour of the utterance 'Op een dag kwam een vreemdeling het dorp binnenwandelen' ('One day a stranger came walking into the village'). In panel a) the original contour and its stylization (solid line) are perceptually identical (close-copy stylization), in panel b) they are perceptually equivalent (conveying the same intonational meaning). In b), pitch movements connect the lower and the upper pitch levels which are connected by a baseline and a topline, respectively (dashed lines, see text).*

One factor which made synthetic pitch contours sound more natural, and which was therefore formalized, turned out to be declination (also e.g. Pierrehumbert, 1980). **Declination** is the downtrend of pitch in the course of an utterance which is normally observed in neutral declarative utterances. This means that, on the average, pitch at the beginning of an utterance is higher than at the end. Low-pitch values within an utterance can often be connected by one single line, which is referred to as the 'lower declination line' or **baseline**, whereas high-pitch values can be connected by an 'upper declination line' or **topline**. The transitions of pitch between low and high pitch levels are the pitch movements. The distance between (abstract) declination lines measured perpendicularly to the horizontal time axis is called **excursion size**.

Earlier research on speech synthesis (e.g. Bolinger, 1958; Isačenko & Schädlich, 1964, 1966) has already shown that the presence of a pitch movement realized on a syllable results in perception of a pitch accent. At other places in the utterance the presence of a pitch movement did not induce an accent. Applying the method of close-copy stylization as described above (De Pijper, 1983), an inventory of accent-lending pitch movements in Dutch was determined. The most frequently used accent-lending pitch movements are: two rises (early and late), one fall (early) and a combination of an early rise and a late fall (called rise-fall). Two special ones, used less frequently and in restricted melodic conditions, are: a combination of a late rise with a very late fall, and a fall which comprises typically half the standard range of a accent-lending fall resulting in a 'call contour' ('t Hart et al., 1990).

Earlier research on prominence (e.g. Pierrehumbert, 1979; Liberman and Pierrehumbert, 1984; Hermes and Van Gestel, 1991; Terken, 1991) has shown that prominence of pitch-accented syllables is very much related to the excursion size of the accent-lending pitch movement. Other research has shown that the timing of accent-lending pitch movements within the accented syllable can also influence the perceived prominence of the syllable, but that this effect is relatively small (e.g. Rump, 1992).

In this study we will try to develop a quantitative model describing which information in the pitch contour listeners use when judging relative prominence of pitch-accented syllables. To that effect, a series of experiments has been run in which listeners adjusted interactively certain properties of the pitch contour.

1.2 OUTLINE OF THIS THESIS

After this first, introductory chapter, we will describe in chapter 2 two sets of experiments in which the relative prominence of two pitch-accented syllables, one in the low register and one in the high register, will be compared. The goal of these experiments is to test the ERB hypothesis of Hermes and Van Gestel (1991) according to which pitch-accented syllables are perceived to be equally prominent when excursion sizes of accent-lending pitch movements are equal when measured in E, but not when measured in Hz or st. The experiments by Hermes and Van Gestel were inspired by some informal observations that, in a corpus of speech data from men and women, excursion sizes in the female, high pitch register were larger than those in the normal, male pitch register when measured in Hz, but that they were smaller when they were measured in st. Therefore, they expected that a scale between these two would fit the data best. One possible candidate was the psychophysically-determined ERB-rate scale.
In the first set of experiments described in chapter 2, the stimuli in low and high registers are resynthesized from a male voice. Formants were kept constant so that the stimuli resynthesized in the high register sound as if spoken by a boy's or (male) falsetto voice. In the second set, the stimuli in the low register are again resynthesized from the male voice, but those in the high register are resynthesized from a female voice. The results of both sets of experiments support the ERB hypothesis.

In chapter 3, prominence experiments are described in which three different kinds of accent-lending pitch movements, a rise, a fall, and a rise-fall, are compared, either in the same or in different registers. On the basis of the outcome of these experiments, a simple quantitative model describing the relationship between pitch-level differences (PLDs) in each of the stimuli and perceived prominence is presented.

In chapter 4, the PLD model presented in chapter 3 is evaluated experimentally. On the basis of these experiments an alternative model is presented which can also account for the results of the experiments presented in chapters 2 and 3. The alternative model is meant to make quantitative predictions, but, in fact, it is qualitative in nature. It assumes the creation by the listeners of an abstract, low-pitched, reference level. The perception of the difference between two pitch levels becomes impossible since one of the pitch levels is abstract. According to this alternative model, listeners base their prominence judgments on the difference between a high pitch level, which is actually present in the stimulus, and the low-pitched reference level which is assumed to be abstract. It is plausible that for building the low-pitch reference level, information can be used provided by the low-pitch levels which are also present in the stimulus and, in addition, by the voice characteristics of the speaker. An experimental evaluation of the model shows that it makes rather good predictions.

The prominence model presented in chapter 4 applies to pitch accents in isolated utterances with one accented syllable; this always had the same position within the utterance. However, in real speech accents can occur early or late in an utterance, and one utterance can have more than one accented syllables. Therefore, in chapter 5 experiments will be reported in which the prominence is compared of two accented syllables within one utterance. The stimuli consist of a meaningful Dutch sentence /A'manda gaat naar 'Malta/ which contains two pitch-accented syllables, /mɑn/ and /mɑl/. The results show a significant influence of the rate of baseline declination on prominence and provide additional evidence for the assumption of an abstract low-pitched reference level against which the heights of high pitch levels are rated. The way in which listeners create a low-pitched reference level is also discussed.

In chapter 6, a functional aspect of the prominence of pitch-accented syllables is studied, i.e. the degree of prominence of pitch accents which is known to influence the perceived focus structure of an utterance. The same utterance was used as in chapter 5 ("Amanda gaat naar Malta"). Different focus conditions were elicited by asking questions to which the utterance should be an appropriate answer. Two kinds of experiments were performed. In the first, listeners adjusted either the value of the pitch-peak maximum on /mɑn/ or that on /mɑl/ while the maximum of the other peak was fixed at three different values. It was found that focus has a strong influence on adjusted peak values, but that values of the fixed peak do not systematically influence those values. In the second part, listeners judged which pitch contours would be appropriate under the various focus conditions. It was found that even relatively large prominence differences do not necessarily influence the perceived focus structure of an utterance. Only when the prominence of the second pitch accent is clearly larger than that of the first, or when the pitch peak on the second syllable is absent, are the two pitch accents explicitly unequal. In that case, only one of the pitch accents is in focus while the other is defocused.

In chapter 7, the results of the experiments presented in this thesis are summarized and conclusions are drawn. The main conclusion is that low and high pitch levels play a different role in lending prominence to pitch-accented syllables. It is hypothesized that low-pitch values are used by the listener for creating a low-pitch reference level against which high-pitch values are rated. When pitch movements are in different utterances, as in the /ma'mama/-experiments, equal heights of high-pitch values seem to be a cue for the perception of equal prominence. If the pitch-accented syllables are in one utterance, the pitch peak maximum on the second pitch-accented syllable is slightly lower than that on the first in order to sound equal in prominence. It is hypothesized that the height of high-pitch values is rated locally on the basis of information about the speaker's pitch range, which may be derived from the spectral build-up of the speech signal or voice characteristics. Finally, the relative prominence of pitch-accented syllables does not directly influence the focus structure of an utterance. Only when there is great deviation from the expected prominence under the neutral-focus condition, will the meaning of the utterance be affected. Smaller deviations may affect the naturalness of a pitch contour, but the extent to which this holds remains to be tested in future experiments.

2 PROMINENCE OF PITCH-ACCENTED SYLLABLES AND THE ERB-RATE SCALE[1]

ABSTRACT

Two experiments will be presented, extending earlier research that addressed the problem of the frequency scale for prominence perception in speech intonation (Hermes and Van Gestel, 1991). Statistical analyses of the new data support the conclusion of the first set of experiments that pitch-accented syllables in different pitch registers are perceived as equally prominent when their associated pitch movements have equal excursion sizes, measured in number of ERBs rather than in Hertz or semitones. In the second experiment, it is shown that this conclusion also holds when the stimuli in the low register are spoken by a male voice and those in the high register by a female voice instead of by a transposed male voice.

2.1 INTRODUCTION

Hermes and Van Gestel (1991) concluded that two identical accent-lending pitch movements lend equal prominence to accented syllables that are resynthesized in different pitch registers if their excursion sizes are equal on the Equivalent-Rectangular-Bandwidth-rate (ERB-rate) scale rather than on the linear Hertz scale or the logarithmic semitone scale. The ERB-rate scale (Patterson, 1976; Glasberg and Moore, 1990) is a psychophysically-defined frequency scale based on measurements of auditory bandwidths, and has E as a unit.

Hermes and Van Gestel performed a prominence-adjustment experiment in which listeners were asked to adjust the excursion size of an accent-lending

[1] Submitted to the Journal of the Acoustical Society of America as H.H. Rump and Dik J. Hermes: "Prominence of pitch-accented syllables and the ERB-rate scale." The data were also presented at the 128th meeting of the Acoustical Society of America, Austin, Texas, as Dik J. Hermes and H.H. Rump: "Control experiments on the frequency scale of speech intonation," Journal of the Acoustical Society of America **96** (1994), 3349.

pitch movement in one register until its prominence matched that lent by the same pitch movement in a different register. The carrier utterance in the high and low registers was a nonsense /ma'mama/ utterance produced by a male speaker. The second syllable was accented. Three kinds of accent-lending pitch movements, a rise, a fall, and a rise-fall, were realized between a fixed lower declination line or baseline and an upper declination line or topline. The baseline and the topline ran parallel in each of the three different frequency scales. The common reference points for the different scales were the fixed end frequencies of the baselines, being 75 Hz in the low register and 180 Hz in the high register. Stimuli were always presented in pairs. The first stimulus in a pair was referred to as the test stimulus. It had a fixed excursion size during any one trial. It was presented in either the high or the low register. The second stimulus, having a variable excursion size, was referred to as the comparison stimulus, and it was presented in the opposite register. When the test stimulus was in the low and the comparison stimulus in the high register, the adjustment was called *upward*, in the opposite case it was called *downward*.

The results showed that listeners adjusted the excursion sizes in the different registers to be equal in E, rather than in Hz or in st. Hermes and Van Gestel have demonstrated convincingly that one musical listening strategy, i.e. making pitch intervals in the low and the high register equal on the semitone scale, would lead to completely different results from those actually obtained. Scrutiny of their experiments revealed, however, that yet another musical listening strategy would lead to roughly the same results as the ones that were obtained. It turned out that, for the various scale types, equal excursion sizes expressed in E also meant that the pitch levels of the toplines of the stimuli in the high register were about twice as high as those of the toplines of the stimuli in the low register. In other words, the low and the high register were accidentally chosen in such a way that the pitch levels of the toplines of the stimuli differed by almost an octave if the excursion sizes of the pitch movements were equal in E. This is illustrated in Fig. 1. Note that the difference between the pitch levels of the baselines was larger than an octave.

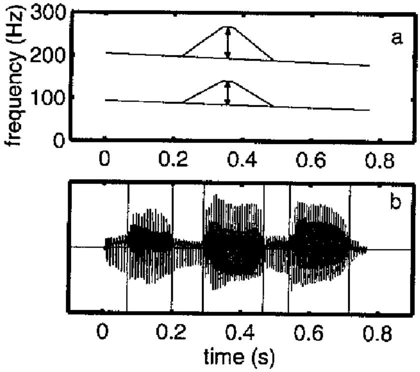


Figure 1. *Illustration of the "octave hypothesis". Pitch contours in the low and the high register containing accent-lending pitch movements with equal excursion sizes expressed in E. Coincidentally, the baselines in the low and the high registers were chosen in such a way that the pitch level of the topline in the high register turned out to be about twice as high as the pitch level of the topline in the low register (octave relationship) when excursion sizes were equal when measured in E, but not when measured in Hz or st.*

According to this so-called "octave hypothesis" listeners attended exclusively to the high-pitch level of the toplines in the low register and to the high-pitch level of the toplines in the high register, adjusting them in such a way that an octave relationship between these high pitch levels resulted. According to Hermes and Van Gestel (1991), listeners adjusted excursion sizes to be equal on the ERB-rate scale. This will be referred to as "E hypothesis". In Table I, the values of the pitch maxima predicted by the octave and E hypotheses are compared to those obtained by Hermes and Van Gestel (1991).

It can be seen in Table I that the values predicted by the octave and E hypotheses are quite close together and that the obtained values are in some cases closer to the values predicted by the octave hypothesis and in others closer to the E hypothesis.

Table I. *Values of the pitch peak maxima (Hz) predicted by the octave (columns 2 and 5) and E hypotheses (columns 3 and 6) and the values obtained by Hermes and Van Gestel (1991) (columns 4 and 7). Column 1 indicates on which scale the excursion-size ranges in the different registers were equal. Columns 2-4 list the values for the upward sessions, columns 5-7 for the downward sessions. The first part of the table lists the values for the rises and the rise-falls, the second part lists the values for the falls.*

RISES / RISE-FALLS

	upward adjustments			**downward** adjustments		
scale	octave hyp.	E hyp.	obtained	octave hyp.	E hyp.	obtained
E	212	227	225 / 232	108	100	106 / 105
	230	238	236 / 242	113	108	111 / 105
	244	250	248 / 240	119	115	114 / 109
	260	263	254 / 263	125	123	119 / 123
	278	270	263 / 265	131	132	133 / 125
	294	286	275 / 268	135	139	135 / 130
st	186	235	225 / 232	116	105	110 / 106
	196	255	236 / 242	125	115	114 / 113
	206	270	248 / 240	135	130	132 / 122
	218	292	254 / 263	147	146	139 / 134
	228	318	263 / 265	157	165	152 / 152
	238	340	275 / 268	166	180	160 / 154
Hz	244	221	232 / 235	102	98	103 / 98
	266	230	253 / 247	107	103	107 / 109
	290	238	263 / 255	111	109	114 / 107
	312	245	267 / 267	113	114	114 / 111
	334	254	279 / 284	119	120	121 / 114
	358	263	296 / 293	122	125	125 / 122

In the case of the rise-fall, the pitch maximum in the second syllable was measured. In the case of the single rise and fall, the high-pitch levels after the rise in the second syllable, and before the fall in the first syllable, were measured.

Table I (continued)

	FALLS					
	upward adjustments			**downward** adjustments		
scale	octave hyp.	E hyp.	obtained	octave hyp.	E hyp.	obtained
E	236	244	242	114	106	105
	254	256	253	119	115	114
	270	270	266	125	123	118
	290	278	263	131	133	127
	308	294	276	139	143	135
	328	313	288	147	152	148
st	202	235	238	128	115	120
	212	241	239	139	128	140
	224	250	240	147	146	153
	236	260	259	161	165	153
	250	270	256	172	184	177
	264	282	267	185	202	175
Hz	274	251	253	106	101	105
	298	263	265	111	108	111
	322	282	265	114	114	114
	344	299	290	119	119	119
	370	308	290	122	126	125
	400	316	294	128	133	127

Earlier experiments on prominence have shown that the variation of the pitch levels of the topline is more relevant for the perception of prominence than the variation of the pitch levels of the baseline (c.f. Pierrehumbert, 1981; Sluijter and Terken, 1993). This was the reason why the prominence lent by the pitch movements was controlled by varying the variable topline and keeping the baseline fixed. This, however, may have biased the subjects in such a way that they attended even more closely to the pitch levels of the topline of each of the stimuli. The octave relationship between these pitch levels may have influenced the outcome of the Hermes and Van Gestel experiments, because it is very easy to perceive, to produce, and, therefore, to adjust octave relations between pitch levels. This octave hypothesis was tested in a new experiment.

In addition to the **octave hypothesis** and the **E hypothesis**, we will test the same hypotheses as Hermes and Van Gestel (1991) did. These are, first, the ''**st hypothesis**'' which proposes that excursion sizes of the pitch movements in the low and high register are equal on the semitone scale. Since the ERB-rate scale is between the linear Hertz scale and the logarithmic semitone scale, equal excursion sizes expressed in st would mean that the excursion sizes of the pitch movement in the high register are larger than those in the low register when measured in E. Second, the ''**Hz hypothesis**'' which proposes that pitch-accented syllables are perceived to be equally prominent when the excursion sizes expressed in Hz are equal. That would mean, as can also be inferred from Fig. 1, that the excursion sizes of the pitch movements in the high register are smaller than those in the low register when measured in E. The four predictions will be tested in experiment 1.

In the second experiment, the hypothesis is tested that the use of a male voice in the high register, which sounded like a male falsetto voice, may have influenced the results of the previous experiments. It has been suggested ('t Hart, personal communication) that the pitch range of a male falsetto voice is much narrower than that of a female voice, so that a given excursion size of the accent-lending pitch movement would lend more prominence to the accented syllable when spoken by a falsetto than by a female voice. The hypothesis that the ('assumed') pitch range of a speaker influences the perceived prominence is tested in experiment 2.

2.2 EXPERIMENT 1: DIFFERENT BASELINES IN THE TWO REGISTERS

The same kind of adjustment experiment was carried out as in Hermes and Van Gestel (1991). In this experiment, however, the pitch contours only comprised a rise-fall on the accented syllable. In order to test whether the octave relationship has affected the outcome of the previous experiment, the baselines in the low and high register were varied systematically. In some instances, the octave relation would lead to equal excursion sizes in E, as had been the case in the previous experiment, in other instances, the octave relation would lead to different excursion sizes of the pitch movement in the low and high registers, when expressed in E (see below). The E hypothesis predicts equal excursion sizes in E in every instance. In addition, the results will be compared to the Hz hypothesis and the st hypothesis.

2.2.1 Procedure

Three different stimulus sets in the low register were compared to five different stimulus sets in the high register. In each stimulus set, the baseline was fixed. The stimulus sets will be referred to by the lowest frequency (in Hz) of their baseline. The lowest frequencies in the low register were 67, 75, and 83 Hz, in the high register they were 140, 160, 180, 200, and 220 Hz. In the experiments by Hermes and Van Gestel, the lowest frequencies had been 75 and 180 Hz. In every instance, the baseline and the topline ran parallel in the ERB-rate frequency domain. In addition, the baselines in the low and the high register ran parallel in E. The rate of declination was 0.70 E/s (about 4.9 st/s). In order to convert Hz values into E values the formula by Glasberg and Moore (1990, p. 114) was used:

Number of ERBs, E = 21.4 * LN(0.00437 * F + 1)/(LN10) (F in Hz).

The octave hypothesis and the E hypothesis coincide more or less in the comparisons of 67 with 160, 75 with 180, and 83 with 200 Hz. The octave hypothesis predicts that when the difference between the low and the high register increases, the excursion sizes expressed in E will be smaller in the high register than in the low register. When the difference between the low and the high register decreases, however, the excursion sizes expressed in E will be larger in the high register than in the low register.

The experiment was divided into six sessions, each lasting about half an hour. There were three upward and three downward sessions (see above). In each session, the baseline of the stimuli in the low register was the same. A session consisted of five blocks in which the baselines of the stimuli in the high register were the same. For example, in one block of a downward session, the test stimulus had a lowest frequency of 140 Hz, while the comparison stimulus had a lowest frequency of 67 Hz.

Each block consisted of six trials in which the excursion size of the test stimulus was fixed at 0.56, 0.83, 1.11, 1.38, 1.67, and 1.94 E. The excursion size of the comparison stimulus could be varied from 0 to 2.5 E (in the low register about 94 Hz or 14 st, in the high register about 126 Hz or 9 st) by selection from a set of prepared stimuli. The range of 2.5 E was divided into 18 equidistant steps of 0.14 E. This size of step was half the step size used by Hermes and Van Gestel. [Although the step size was still much larger than the difference limen for the perception of tonal changes (e.g. Rossi, 1971, 1978), it was much smaller than the differences between excursion sizes which were suggested to be relevant in the perception of prominence in speech, being about two to three semitones ('t Hart, 1981)]. The initial excursion size of the comparison stimulus was zero, as was the case in the previous experiments. The total number of adjustments performed by each subject was 180 (6 sessions x 5 blocks x 6 trials).

Six subjects took part in all three parts of the experiment. They were all staff members of the institute. Some of them had participated in the experiments by Hermes and Van Gestel. None of the subjects reported any hearing deficiencies.

2.2.2 Results and discussion

One of the subjects did not follow the instructions, so that his data were discarded from further analysis. The results, averaged across the remaining five subjects, are displayed in Fig. 2. Along the horizontal and the vertical axes, the excursion sizes in E of the stimulus in the low and the high registers are indicated respectively. Each panel shows the combined results of an upward session, indicated by black circles and vertical standard deviation bars, and a downward session, indicated by open circles and horizontal standard deviation bars. The standard deviation bars indicate the combined effect of variability between and within subjects.

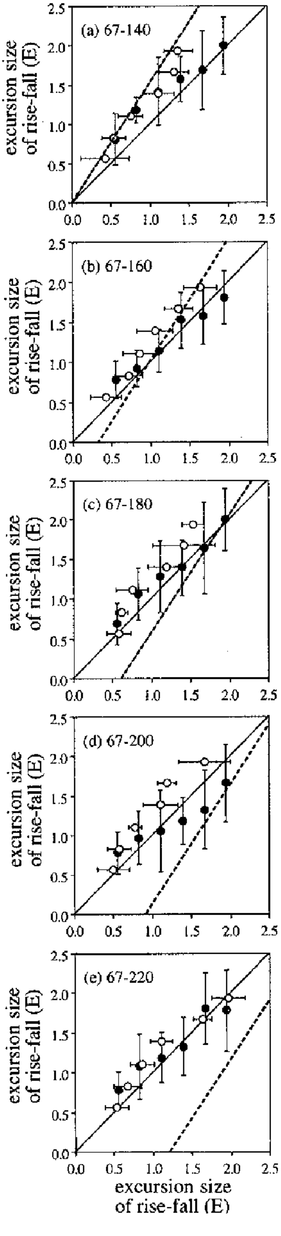
The lowest frequency in the low register was 67 Hz in the panels (a), (b), (c), (d), and (e), it was 75 Hz in the panels (f), (g), (h), (i), and (j), and it was 83 Hz in the panels (k), (l), (m), (n), and (o). The lowest frequency in the high register was 140 Hz in the panels (a), (f), and (k), 160 Hz in the panels (b), (g), and (l), 180 Hz in the panels (c), (h), and (m), 200 Hz in the panels (d), (i), and (n), and 220 Hz in the panels (e), (j), and (o). The E hypothesis (solid line) and the octave hypothesis (dotted line) make roughly the same predictions in the panels (b), (h), and (n).

In order to test the hypotheses, the absolute distances between observed and predicted values were determined[2]. Distances were measured perpendicularly to the horizontal axis and expressed in E. Table II lists for each subject, collapsed over all register differences, the average absolute distance of the observed values from the E hypothesis, represented by the diagonal in Fig. 2. The average absolute distance between the observed values and the octave hypothesis, the Hz hypothesis, and the st hypothesis are also shown. In addition, the means of these averages are displayed.

In predicting the st, Hz, and octave values, corrections were made for the range of excursion sizes, i.e. the minimum predicted value was set to 0 E and the maximum predicted value was set to 2.5 E. It was not necessary, of course, to make such corrections when predicting the ERB values. The corrections had as a result that the distances for the octave, semitone, and Hertz predictions became slightly smaller.

[2] Another method of testing, which we did not choose to use but which would have led to more or less the same results, would have been to determine mean *real* distances and to compare their standard deviations.

Figure 2. *Results of experiment 1. Mean adjusted excursion sizes of the pitch movements in the high register (black dots, vertical standard deviation bars) and the low register (open dots and horizontal standard deviation bars). In each panel, the combined results of an upward and a downward session are shown: (a) 67-140 Hz, (b) 67-160 Hz, (c) 67-180 Hz, (d) 67-200 Hz, (e) 67-220 Hz, (f) 75-140 Hz, (g) 75-160 Hz, (h) 75-180 Hz, (i) 75-200 Hz, (j) 75-220 Hz, (k) 83-140 Hz, (l) 83-160 Hz, (m) 83-180 Hz, (n) 83-200 Hz, (o) 83-220 Hz. The diagonal indicates equal excursion sizes expressed in E, the dashed line indicates an octave relationship between pitch levels of the toplines in the low and high registers.*



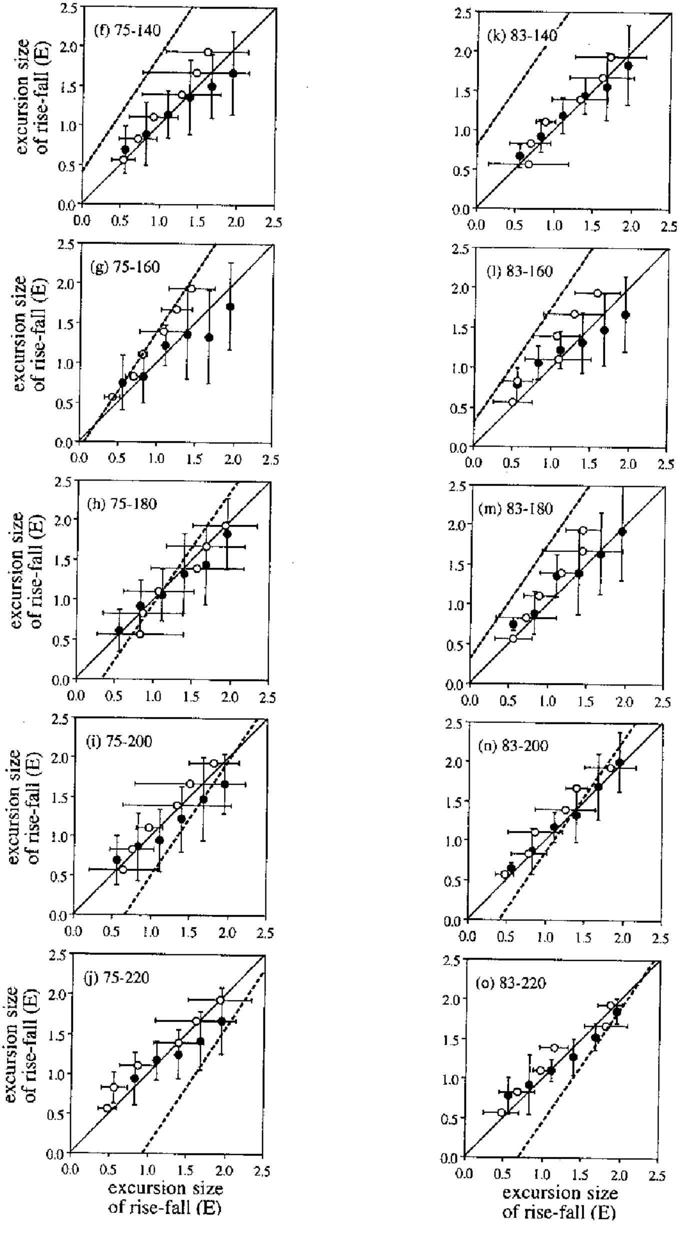

(f) 75-140
(g) 75-160
(h) 75-180
(i) 75-200
(j) 75-220
(k) 83-140
(l) 83-160
(m) 83-180
(n) 83-200
(o) 83-220
excursion size
of rise-fall (E)
0.0
0.5
1.0
1.5
2.0
2.5
excursion size
of rise-fall (E)

Table II. *Average absolute distances expressed in E between the observed values and those predicted by the different hypotheses, collapsed across registers. Column 1: subject; columns 2-5: E hypothesis, octave hypothesis, st hypothesis, and Hz hypothesis.*

	Hypothesis			
	E	oct	st	Hz
Subj.				
JT	0.303	0.538	0.323	0.606
JP	0.370	0.663	0.743	0.404
LB	0.244	0.453	0.350	0.555
MS	0.326	0.487	0.606	0.407
RS	0.223	0.441	0.397	0.516
Mean	0.293	0.516	0.484	0.498

As can be seen in Table II, the average absolute distance from the E hypothesis was smallest and the average absolute distance from the octave hypothesis was largest. The absolute distance between observed values and the st hypothesis was smaller than that between observed values and the Hz hypothesis, but both were larger than the absolute distance between observed values and the E hypothesis. These trends were found in the data of almost every individual subject (cf. Table II). A repeated-measures ANOVA across subjects showed that the main effect of Hypothesis was significant ($F_{(3,12)}$ = 4.20, $p < 0.030$). (Planned) comparisons between the average absolute distance from the E hypothesis and the distances from the other hypotheses revealed that all these differences were significant (ERB vs. st: $F_{(1,12)} = 7.09$, $p < 0.05$; ERB vs. Hz: ($F_{(1,12)} = 8.17$, $p < 0.05$; ERB vs. octave: $F_{(1,12)} = 9.67$, $p < 0.01$). A new repeated-measures ANOVA, leaving out the E hypothesis, showed that the differences between the distances from the st hypothesis, from the Hz hypothesis, and from the octave hypothesis were non-significant ($F_{(2,8)} = 0.07$, $p < 0.93$). Notice that the fact that the predictions of the Hz hypothesis, the st hypothesis, and the octave hypothesis do not differ significantly does not imply that they make the same predictions. Only the amount of the deviation from the observed values is equal.

Since the observed values were closest to those predicted by the E hypothesis, we then calculated the average distance between them. The analysis showed that this distance was 0.093 E (equal to about 3 Hz or 0.7 st in the low register and 4 Hz or 0.4 st in the high register), which means that the excursion sizes of the pitch movement in the high register were, on average, about 0.1 E larger than those of the pitch movement in the low register when lending equal prominence. Though small, this difference turned out to be significantly different from zero (T = 7.76, df = 899, $p < 0.001$).

The results indicate that the E hypothesis was to be preferred over the st and the Hz hypothesis. It was also shown that the octave hypothesis was clearly not supported by the new data. Rather, the listeners perceived the pitch-accented syllables as equally prominent when the excursion sizes of the pitch movement in the low and the high registers were equal when expressed in E.

The excursion sizes of the pitch movement in the high register turned out to be on average about 0.1 E larger than those in the low register. Although this difference turned out to be highly significant, it was very small and probably negligible. Indeed, according to 't Hart (1981), such small differences between excursion sizes are very likely to go unnoticed. This conclusion is supported by the subjects' remarks that they did not hear any differences between comparison stimuli being apart one step of 0.14 E in the excursion size continua.

In summary, the results of experiment 1 suggest that identical pitch movements in different registers lend equal prominence to the accented syllables if the changes of pitch comprise an equal number of ERBs. In conclusion, the reanalysis of the previous experiments by Hermes and Van Gestel has shown that their stimuli were accidentally chosen in such a way that the outcome may have been influenced by the octave effect. It has now been shown that this octave effect was not crucial. It was also shown that the Hz and the st hypotheses make predictions that are statistically significantly different from the obtained results. The main conclusion of Hermes and Van Gestel therefore still stands, i.e. that identical pitch movements resynthesized in different registers lend equal prominence when their excursion sizes are equal when expressed in E.

As can be seen from the plots in Fig. 2, the data for the upward and the downward session differ slightly. If the test stimulus was in the low register, the excursion size of the comparison stimulus in the high register was likely to be adjusted somewhat smaller than that of the test stimulus, expressed in E. If the test stimulus was in the high register and the comparison stimulus was in the low register, this effect was not found. Hermes and Rump (1994) hypothesized that this was due to the fact that the subjects were reluctant to adjust large excursion sizes for the pitch movement in the high register. It may also indicate an order effect, because the comparison stimulus was always the second stimulus of a stimulus pair. Yet another possible explanation, which will be tested in Experiment 2, relates to the fact that the stimuli in both the low and high register were resynthesized from the same utterance which was spoken by a male person.

2.3 EXPERIMENT 2: MALE AND FEMALE VOICES

In the previous experiments, the stimuli in the low and high register were resynthesized from an utterance which was produced by a male speaker. In order to obtain the high-pitched stimuli, the original low pitch was transposed to a high register more than an octave above the low register. The resulting high-pitched stimuli sounded as if spoken in a boy's or a male falsetto voice. According to 't Hart (personal communication) this may have influenced the results of the previous experiments since the natural pitch range of such high-pitched voices may be smaller than the pitch range of an original, high-pitched female voice. 't Hart hypothesized that the adjusted excursion sizes in the high register were therefore found to be somewhat smaller than when a female voice would have been used. As a result, the ERB-rate-scale preference was found instead of a semitone-scale preference. In a different experimental set-up in which listeners had to rate the liveliness of different speakers' voices, Traunmüller and Eriksson (1995) found indications that, among other factors, the sex of the speaker influenced the liveliness ratings: a male and a female voice were rated to be about equally lively when excursion sizes were more or less equal when expressed in st. This indicates that, at least in their experiments, listeners' expectations about pitch ranges may indeed have played a role.

It is not clear from the literature whether a male or a female voice has a wider pitch range, or for that matter whether pitch ranges are equal in either of the different frequency scales (for an overview see Henton, 1989). Therefore, it is also possible that the excursion sizes of the female voice will be found to be smaller than those of the male voice when lending equal prominence. In this experiment we will test the assumption made by 't Hart, i.e. that a semitone-scale preference will be found if in the high register a female voice is used instead of a transposed male voice. As mentioned before, the st hypothesis predicts that excursion sizes in the high register are larger than those in the low register when expressed in E. The results will also be compared to the predictions made by the Hz hypothesis.

2.3.1 Procedure

As in the previous experiments, the carrier utterance was again the nonsense /ma'mama/ utterance with a pitch accent on the second syllable. The stimuli in the low register were resynthesized from the male voice, whereas the stimuli in the high register were resynthesized from a female voice. The temporal structures of the male and the female utterance were made identical by making the number of samples per speech segment equal. As a result, some parts of the female utterance were shortened somewhat. This was done because Hermes and Rump (1994) have found indications that the time interval between the vowel onsets of the accented and the previous syllable might influence the perceived prominence.

The experimental set-up was almost identical to the one in experiment 1. The only difference was that in the present experiment, the stimuli in the low register only comprised those with a fixed end frequency of the baseline of 75 Hz. In addition, the stimuli in the high register, comprising the five end frequencies (140, 160, 180, 200, or 220 Hz) which were used in experiment 1, were now resynthesized with a female voice.

The experiment was again divided into an upward and a downward session, each lasting about half an hour. As before, in each session, there were five blocks in which the end frequency of the high register was kept constant. In each block, there were six trials in which the test stimulus had a fixed excursion size of 0.56, 0.83, 1.11, 1.38, 1.67, or 1.94 E. The excursion size of the comparison stimulus could be varied from 0 to 2.5 E (in the low register about 94 Hz or 14 st, in the high register about 126 Hz or 9 st) by selection of prepared files from the appropriate stimulus sets. The range was divided into 18 equidistant steps of 0.14 E. The initial excursion size of the comparison stimulus was zero, as was the case in the previous experiments. The total number of adjustments performed by each subject was 60 (2 sessions x 5 blocks x 6 trials).

Seven subjects took part in the experiment. They were all staff members of the institute. Some of them had taken part in the experiments by Hermes and Van Gestel and most of them in the experiments by Rump and Hermes. None of the subjects reported any hearing deficiencies.

2.3.2 Results and discussion

The results averaged across subjects are displayed in Fig. 3. Along the horizontal and the vertical axes, the excursion size in E of the stimulus in the low and the high register is indicated respectively. Each panel shows the combined results of an upward session, indicated by black circles and vertical standard deviation bars, and a downward session, indicated by open circles and horizontal standard deviation bars, in which the low and the high register were kept constant. The standard deviation bars indicate the combined effect of variability between and within subjects. The lowest frequency of the baseline in the low register was 75 Hz in all panels. The lowest frequency in the high register was 140 Hz in panel (a), 160 Hz in panel (b), 180 Hz in panel (c), 200 Hz in panel (d), and 220 Hz in panel (e). The solid line (diagonal) indicates the predictions by the E hypothesis, the dotted line indicates those by the st hypothesis, and the dashed line indicates those by the Hz hypothesis.

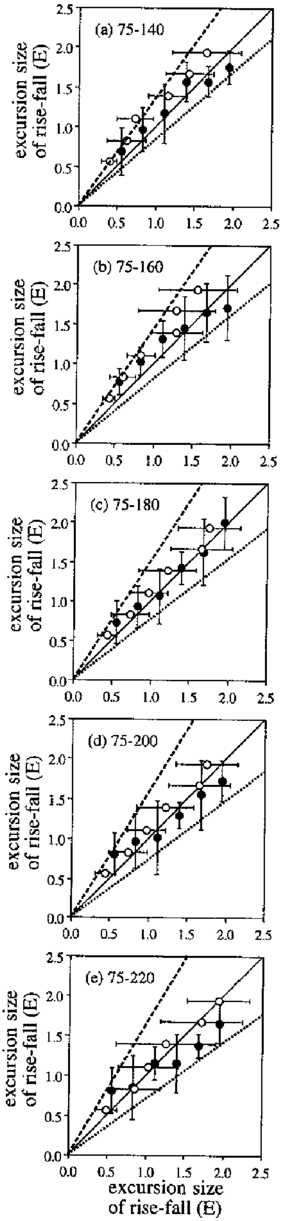


Figure 3. *Results of experiment 2. Mean adjusted excursion sizes of the pitch movements in the high register (black dots, vertical standard deviation bars) and in the low register (open dots, horizontal standard deviation bars). In each panel, the results are shown for an upward session and a downward session in which the baselines in the low and high register were kept constant: (a) 75-140 Hz, (b) 75-160 Hz, (c) 75-180 Hz, (d) 75-200 Hz, (e) 75-220 Hz. The diagonal indicates equal excursion sizes in the low and high register expressed in E. The dotted line indicates equal excursion sizes expressed in st, the dashed line indicates equal excursion sizes expressed in Hz.*

Table III lists for each subject the average of the absolute distance between the observed values and those predicted by the E hypothesis, which is represented by the diagonal in Fig. 3. Distances were again measured perpendicularly to the horizontal axis and expressed in E. Also shown are the average absolute distances between the observed values and those predicted by the st hypothesis (the dotted line in Fig. 3), and between the observed values and those predicted by the Hz hypothesis (the dashed line in Fig. 3). Averages are across registers. In addition, the means, averaged across subjects, are displayed.

Table III. *As in Table II, average absolute distances expressed in E between the observed values and those predicted by the different hypotheses, collapsed across registers. Column 1: subject; columns 2-4: E hypothesis, st hypothesis, and Hz hypothesis. The last row presents the means averaged across subjects.*

	Hypothesis		
	E	st	Hz
Subj.			
AS	0.322	0.663	0.379
JT	0.280	0.358	0.562
JP	0.313	0.391	0.576
LB	0.241	0.322	0.564
MM	0.352	0.484	0.512
MF	0.292	0.784	0.166
RS	0.169	0.437	0.452
Mean	0.281	0.491	0.459

As can be seen in Table III, the average absolute distances from the E hypothesis were smallest for almost every individual subject. The across-subjects mean for the Hz hypothesis was slightly smaller than that for the st hypothesis, but they were both larger than that for the E hypothesis. A repeated-measures ANOVA across subjects showed that the main effect of

Hypothesis tended to be significant ($F_{(2,12)} = 3.56$, $p < 0.061$), which indicates that the prediction of the E hypothesis was better than those of the st and the Hz hypothesis.

The average distance in E between the observed values and those predicted by the E hypothesis was calculated to be 0.090 E, which is about 3 Hz or 0.7 st in the low register and about 4 Hz or 0.4 st in the high register. This means that the excursion sizes of the pitch movements in the high register were on average about 0.1 E larger than those of the pitch movements in the low register when lending equal prominence. It is worthwhile to remark that this difference turned out to be significantly different from zero ($T = 5.40$, $df = 419$, $p < 0.001$), although it may be perceptually negligible.

2.4 GENERAL DISCUSSION

The results show that the E hypothesis is to be preferred over the st and the Hz hypothesis when expressing excursion sizes of pitch movements that lend equal prominence. The results by Hermes and van Gestel (1990) and those of Experiment 1 were replicated in Experiment 2, but this time with the stimuli in the low and the high register spoken by a male and a female voice, instead of spoken by just a male voice. This means that pitch-accented syllables are perceived to be equally prominent when pitch movements in the low and high register have equal excursion sizes expressed in E, but not in Hz or st, and that this is independent of the fact of whether the stimuli in the high register are resynthesized from an original high-pitched female utterance or from a high-pitched transposed-male utterance. Against the expectations by 't Hart and unlike in the experiments by Traunmüller and Eriksson (1995), listeners' expectations about pitch ranges did not play a role. This implies that our results are proof against factors that are not directly related to prominence. The octave relationship, too, turned out to be of no importance.

There are various reasons why the results obtained in the present study differ from those obtained by Traunmüller and Eriksson (1995). First, it might be that "liveliness" estimations are judged on a different frequency scale than prominence estimations. Since this seems rather unlikely, we will further ignore this possibility. Second, in the context of liveliness estimations, the use of the falsetto voice register is not neutral so that the listeners' impression of liveliness is likely to be influenced in the sense that they may systematically have over- or underestimated the liveliness of the falsetto voice, just as they overestimated the liveliness of the ''child's voice'' under the original-duration condition (Traunmüller and Eriksson, 1995, p. 1910). Since prominence estimations are more likely to be based on pitch values in the accented syllable relative to those in neighbouring syllables (see Hermes and Rump, 1994), such a bias is less likely to be introduced in prominence estimation. Third, the ''male'' voice used by Traunmüller and Eriksson (1995) was resynthesized by lowering the formants of the original female utterance. Changing the formants will have an impact on the perception of liveliness: If lowering the formants decreases the liveliness of a voice, this procedure will have decreased the liveliness estimations of the male voice. This, too, may have introduced a bias toward the semitone scale. Again, prominence estimations are less likely to suffer from such a bias, for the same reason as mentioned before.

Then, there are two experimental reasons which may have biased the results by Traunmüller and Eriksson toward the logarithmic frequency scale. First, for each register, they used seven LPC-resynthesized versions of an utterance spoken originally by a female speaker. The ranges of the two registers covered by these two sets of stimuli were equal on a logarithmic scale. This is motivated by stating that the natural range of male and female voices are equal in semitones ''since the F_0 excursions observed in the most lively types of discourse are actually larger in the speech of women than in that of men, even if expressed in semitones'' (p. 1908). Actually, in neutral speech the pitch range of women is smaller than that of men when expressed in semitones (Graddol and Swann, 1983; Henton, 1989), which favours the ERB-rate scale. (The fact that women have a larger pitch range in the ''most lively types of discourse'' may be attributed to the fact that, in this situation, the use of the falsetto register is socially accepted for females but not for males.) This use of ranges equal on a semitone scale for the male and the female voices in itself may have introduced a bias toward the logarithmic frequency scale.
Second, the subjects may not have rated the liveliness on the basis of a direct comparison with the standard, but relative to the two ranges of the registers of

the two sets of stimuli. Moreover, within each register there were only seven different stimuli with standard deviations rather far apart. The stimulus with the lowest liveliness had a standard deviation of about half a semitone. The next stimulus had a standard deviation of one and a half semitone. Perceptually, this is very wide apart, so the subjects will have easily learned to recognize the two stimuli in the two registers with the lowest standard deviation. The other two stimuli that the subjects may have learned to recognize are the middle ones within a register. This may be enhanced by the use of the stimulus with the original pitch contour as the standard stimulus. It is well known that LPC artifacts become more audible as the pitch of the resynthesized contour is further away from the original. In this respect, it is remarkable that in Fig. 2b of Traunmüller and Eriksson (1995, p. 1909) the first and the middle points of the two graphs almost completely coincide. The other five points are higher for the female voice than for the male voice, which is consistent with the proposition that intonation is perceived on an ERB-rate scale.

The results of our Experiments 1 and 2 are also consistent with respect to the finding that there was a small difference of about 0.1 E between the excursion sizes of the pitch movement in the low and high registers. The excursion sizes of the pitch movements in the high register were on average about 0.1 E larger than those in the low register. Although this difference turned out to be significant, it was very small, and, as mentioned before, is very likely to go unnoticed, since it is considerably smaller than excursion-size differences of about 2 to 3 st which are supposed to be perceptually relevant ('t Hart, 1981).
In conclusion, these experiments provide statistical evidence in support of the conclusion by Hermes and Van Gestel (1991) that pitch-accented syllables are perceived as equally prominent when their excursion sizes in the low and high registers are equal in E rather than in Hz or st. The absolute differences between the various scales, however, are relatively small, which may be the reason why, under different, less controlled experimental conditions others may have found that either the Hertz or the semitone scales are more suited than the ERB-rate scale to expressing equality of excursion sizes (see Henton, 1989).

3 PROMINENCE IN SPEECH INTONATION INDUCED BY RISING AND FALLING PITCH MOVEMENTS[3]

ABSTRACT

The object of this study was to investigate whether subjects are able to compare the prominence caused by different types of accent-lending pitch movements and, if so, whether some pitch movements lend more prominence to a syllable than others. These experiments were carried out with the utterance /ma'mama/, with the second syllable accented by either a rise, a fall or a rise-fall. Subjects adjusted the variable excursion size of a comparison stimulus to the fixed excursion size of a test stimulus in such a way that the accented syllable in test and comparison stimuli had equal prominence. The rise-fall was only presented in one ''standard'' position, while the fall and the rise were tested for five different positions in the syllable. Subjects were found to be quite capable of equating the relative prominence of syllables accented by the following types of pitch movement: the rise-fall in standard position, the rise starting before the vowel onset, and the fall whatever its position in the syllable. When lending equal prominence, the early starting rise and the rise-fall had equal excursion sizes. The fall, however, appeared to lend more prominence to a syllable than the rise or the rise-fall of equal excursion size, independent of its position in the syllable. This difference between the fall on the one hand and the rise and the rise-fall on the other increased with increasing declination of the pitch contour. A model is presented which can explain these phenomena quantitatively.

[3] Published as Dik J. Hermes and H.H. Rump: ''Prominence in speech intonation induced by rising and falling pitch movements,'' Journal of the Acoustical Society of America **96** (1994), 83-92. Parts of the data were presented at the 123rd meeting of the Acoustical Society of America, Salt Lake City, Utah, May 1992, as Dik J. Hermes and H.H. Rump: ''Prominence caused by rising and falling pitch movements with different positions in the syllable,'' Journal of the Acoustical Society of America **91** (1992), 2340, and at the ESCA Workshop on Prosody, Lund, as Dik J. Hermes and H.H. Rump: ''The role of pitch in lending prominence to syllables,'' Working Papers **41** (1993), Lund Univ. Press, Lund, pp. 28-31.

3.1 INTRODUCTION

It is well known that the course of pitch plays an important role in lending prominence to accented syllables (Fry, 1958). Little is known, however, about how this effect comes about. The idea underlying this study is that we can get to know something about this by comparing the prominence induced by different types of pitch movements. In Dutch intonation, at least ten pitch movements can be distinguished ('t Hart, *et al.*, 1990). The distinctions are based on whether the movement is a rise or a fall, on whether it extends over one or over more than one syllable, on its position in the syllable, and on its excursion size. These pitch movements are realised between declination lines, declination being defined as the gradual lowering of the pitch in the course of an utterance or part of an utterance. In the description of Dutch intonation, two declination lines are distinguished, an upper and a lower declination line.

Of the ten established pitch movements, at least three, two rises and one fall, accentuate the syllable in which they occur ('t Hart *et al.*, 1990). The rises are distinguished by their temporal positions in the syllable, one starting early in the syllable, the other late. Both have an excursion size comprising the whole distance between the lower and the upper declination lines. The accent-lending fall starts before the vowel onset, but later than the early rise. It has a normal excursion size comprising the whole distance between the upper and the lower declination lines. Another type of fall, not used in the experiments described here, can also accentuate a syllable, but this depends on its position in an intonational phrase. It differs by its later timing and by its smaller excursion size, as it does not reach the lower declination line. A fifth, very common, way to accent a syllable is to supply it with a combination of an early rise and a fall. This pitch configuration will be referred to as rise-fall. In neutral speech, the accent-lending pitch movements occurring most often are the early rise, the rise-fall and the fall.

The experiments described here were prompted by the informal observation that the succession of an accent-lending rise and an accent-lending fall on the next accented syllable can often be replaced by two rise-falls without affecting the intonational meaning of the sentence. This indicates that the prominence of these syllables, though lent by different pitch movements, is comparable. This issue is investigated here by having subjects compare the prominence of the accented syllables of two utterances which differ only in that the prominence is lent by two different pitch movements.

The question addressed in the first experiment was whether such comparisons could be carried out reliably and, if so, how the excursion sizes of these pitch movements relate to each other when lending equal prominence. The second experiment was carried out to investigate the extent to which the prominence of a syllable depends on the temporal position of the pitch movement in the syllable. The results of these experiments led to the formulation of a model, which was tested in the third experiment.

3.2 EXPERIMENT 1

From the results of previous experiments (Hermes and Van Gestel, 1991), it was concluded that pitch movements in speech intonation are best expressed on a psycho-acoustically defined frequency scale, the ERB-rate scale (Patterson, 1976; Glasberg and Moore, 1990) with E, number of ERBs, as unit. As a by-product of these investigations, it was found that subjects are quite capable to compare the relative prominence of accented syllables in utterances presented in different pitch registers. In the present experiment, subjects were asked to compare the relative prominence of syllables accented by different types of pitch movement: a rise, a rise-fall or a fall. In this first experiment, the temporal positions of the pitch movements were fixed and had standard values as used in the IPO text-to-speech system (Collier, 1991).

3.2.1 Method

The stimuli consisted of modified versions of one utterance /ma'mama/ produced by a male speaker. Its duration was 0.77 s. The vowel onsets occurred 0.08, 0.30 and 0.55 s after the start of the utterance. The second syllable carried the accent. Pitch modifications were applied with the pitch-synchronous overlap and add (PSOLA) technique (Hamon *et al.*, 1989), resulting in high-quality speech stimuli. Duration and amplitude relations were kept constant. The stimuli were resynthesized with a rise, a rise-fall or a fall, realised between declination lines which run parallel on an ERB-rate scale as shown in Fig. 1. The lower declination line of the stimuli was fixed. It started at 3.17 E (93 Hz), and ended at 2.63 E (75 Hz), which corresponded to 0.70 E/s (about 4.9 semitones/s). The prominence of the accented syllable was controlled by the excursion size of the pitch movement, i.e. the vertical distance between the lower and the upper declination lines.

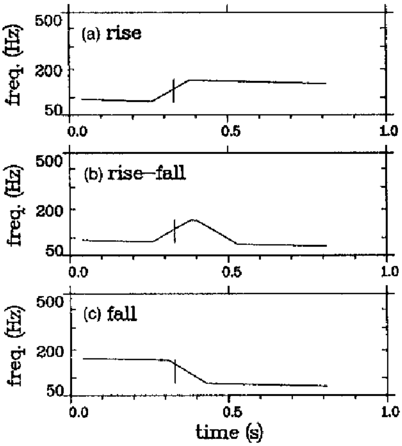


Figure 1. *The three contours with the pitch movements used in experiment 1. The vertical bars represent the vowel onset. Equal intervals on the vertical frequency axis represent equal distances on an ERB-rate scale.*

In all sessions the method of adjustment was used. In each session two stimuli with different pitch movements were repeatedly presented to the subject with an interstimulus interval of a few hundred ms[4]. This rather long interval guaranteed that subjects perceived the stimuli as two utterances, each with one accented syllable in an identical position in the utterance, and not as one utterance with two accented syllables. This was done because the relation between excursion size and prominence in two accented syllables within one utterance is quite complicated (Terken, 1991).

The first stimulus of each trial in an adjustment run will be referred to as the

[4] This interval could vary up to about one second, as the computer operating system interfered with the exact timing of this experiment.

test stimulus. It was fixed within one run. The second stimulus, referred to as the comparison stimulus, had a variable excursion size. In the first trial the excursion size in the comparison stimulus was zero. Subjects were asked first to increase the excursion size in the comparison stimulus so that the prominence of its accented syllable clearly exceeded the prominence of the accented syllable in the test stimulus. In the succeeding trials, the subjects were asked to decrease and increase the excursion size in the comparison stimulus until it was judged to give the same prominence as the pitch excursion in the test stimulus. When the subject had done this, the next run started. Each session comprised six runs, one for each of the six different excursion sizes in the test stimulus: 0.56, 0.83, 1.11, 1.38, 1.67 and 1.94 E. (In this register 1 E corresponds to about 6.1 semitones.) The six different excursion sizes in the test stimulus were presented with a different random order in each session. Figure 2 shows the experimental set-up of a session in which the prominence of a syllable accented by a rise-fall is adjusted to the prominence of a syllable accented by a fall. The range of the six test stimuli used within one session is shown at the left-hand side. The range of the adjustable comparison stimuli, 2.5 E, is shown at the right-hand side. The step size between the stimuli was 0.278 E, resulting in a total of ten possible comparison stimuli. In all these experiments, test and comparison stimuli contained different kinds of pitch movements. If the test stimulus in one session contained one kind of pitch movement and the comparison stimulus the other, the session was repeated with the two pitch movements in test and comparison stimulus interchanged. This was done to test the consistency of the subjects, as described in Hermes and Van Gestel (1991). As there were three different pitch movements tested, this resulted in six different sessions in one such experiment.

In the experimental situation, shown in Fig. 2(a), the test stimulus and the comparison stimulus are presented in the same low register. In this situation the subjects might not equate the prominence of the accented syllables, but simply compare average pitch levels in syllables on the upper declination line. Thus the subject might adjust the higher pitch level in the rise-fall of the comparison stimulus in such a way that it would be equal to the average pitch level in the first syllable of the test stimulus. In order to prevent subjects from following such a strategy, a condition was included in which the stimuli were resynthesized in different registers. One such experimental set-up is presented in Fig. 2(b). Here, the test stimulus was synthesized in the original low register with a *(lower)* declination line ending at 75 Hz, while the comparison stimulus was synthesized in a high register with a lower declination line ending at 180 Hz. The same experiment was repeated with the test stimulus in the high register and the comparison stimulus in the low register. For the sake of completeness, the experiment was also carried out with both stimuli in the high register. So, each experiment was carried out in four conditions which each comprised six sessions, yielding a total of 24 sessions. Six subjects participated in these experiments, all of whom had previously shown that they attained the high level of consistency described in Hermes and Van Gestel (1991).

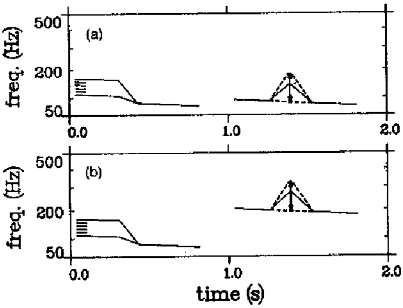


Figure 2. *Setup of an adjustment run. The subject was first presented with one of the six test stimuli displayed on the left-hand side of (a). Then the subject heard the comparison stimulus with an adjustable excursion size shown on the right-hand side. The run stopped when the subject indicated that the same prominence was perceived in the accented syllables of both stimuli. In (b) the same experimental setup is shown for the condition in which test and comparison stimulus were presented in different registers. The interstimulus interval was longer than shown in this figure (see text).*

3.2.2 Results

The way in which the results are presented is shown in Fig. 3. The results of two sessions are combined in one picture. In one of these sessions the pitch movement indicated on the horizontal axis served as test stimulus, while that on the vertical axis served as comparison stimulus. The data points of these sessions are indicated by closed circles. In the other session the situation was reversed: The pitch movement indicated on the vertical axis served as test stimulus, while that on the horizontal axis served as comparison stimulus.

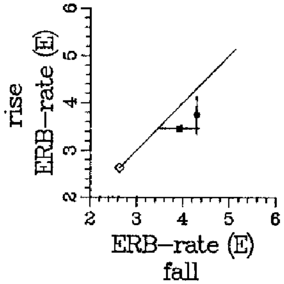


Figure 3. *Diagram showing the results of one run in which the fall was the test stimulus and the rise the comparison stimulus (closed circle), and of one run in which the rise was the test stimulus and the fall the comparison stimulus (closed square).*

The data points of these sessions are indicated by closed squares. The data points indicate averages across subjects. The vertical and horizontal bars represent the standard deviations of data points. The projection of a data point on a coordinate represents the end point of the upper declination line of the stimulus. The diamond shows the end point of the lower declination line of the stimuli (2.6 E, 75 Hz)[5]. This means that the interval on a coordinate between

[5] In the case of a fall or a rise-fall, the endpoint of the upper declination line is virtual, since the pitch contour ends on the lower declination line. In the case of a rise, the endpoint of the lower declination line is virtual, since the pitch contour of a single rise ends on the upper declination line.

the projection of a data point and the diamond represents the excursion size of the stimulus. The diagonal line starting from the diamond shows the excursion sizes of the pitch movements if subjects attributed equal prominence to pitch movements of equal excursion size. If a data point lies above the line, subjects attributed equal prominence when the pitch movement indicated on the vertical axis had a larger excursion size than the pitch movement indicated on the horizontal axis. If a data point lies below the line, subjects attributed equal prominence when the pitch movement indicated on the vertical axis had a smaller excursion size than the pitch movement indicated on the horizontal axis.

The average results of the six subjects are presented in Fig. 4. The comparisons of the rise-fall with the rise, both in the low register, are given in Fig. 4(a). They show that subjects provided the comparison stimulus with about the same excursion size as the test stimulus. In other words, excursion sizes being equal, subjects judged the prominence of a rise and a rise-fall to be equal[6].

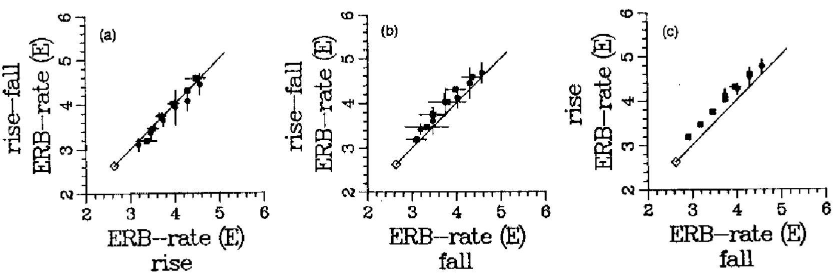


Figure 4 (a)-(c). *See for legend Figure 4 (d)-(l).*

[6] In fact, there appeared to be a small but statistically significant ($p<0.01$) tendency of the data points to lie below the diagonal. (The average distance of the data points from the diagonal was compared with the standard deviation of the mean of this distance.) This result could not be reproduced, however, and the result of an identical experiment is presented in Fig. 6(a). This tendency was also absent in Fig. 4(j) which shows the results of the same experiment carried out with test and comparison stimuli in the high register. It is concluded that this statistically significant deviation is an example of the vagaries of statistics, whereby results will deviate from expectations in a small but real number of cases.

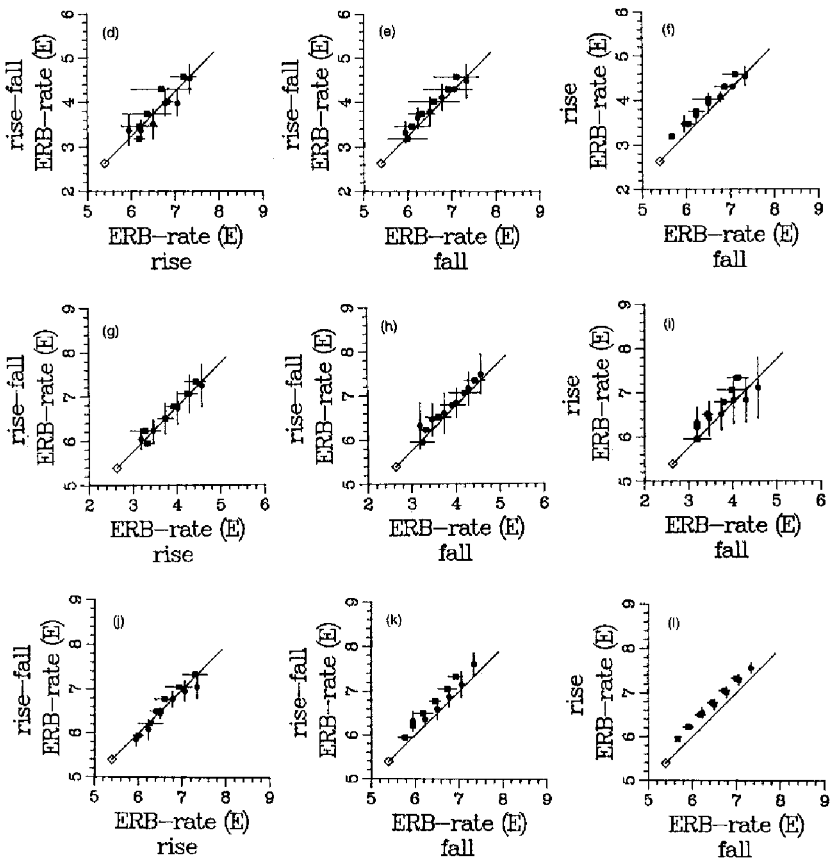


Figure 4 (d)-(l). *Results of the six subjects who participated in experiment 1. In (a), (b), and (c), the results are shown for the sessions in which both test and comparison stimulus were presented in the low register. In (d), (e), (f), (g), (h), and (i), the results are shown for the sessions in which test and comparison stimulus were presented in different registers. In (j), (k), and (l), the results are shown for the sessions in which both test and comparison stimulus were presented in the high register. In (a), (d), (g), and (j) the average adjustments for the rise-fall and the rise are shown, while (b), (e), (h), and (k) show those for the rise-fall and the fall, and (c), (f), (i), and (l) those for the rise and the fall.*

A different result was obtained for the rise-fall compared with the fall. As can be seen in Fig. 4(b), the rise-fall had a larger excursion size than the fall when lending equal prominence. The same result was obtained for the rise compared with the fall, as shown in Fig. 4(c), where the variance was so slight that in some points the standard deviation bars disappear in the data points. The results of this experiment with the test and comparison stimulus in different registers are presented for all six possible combinations in Fig. 4(d)-(i). Though the standard deviations were larger, indicating that the subjects found the task more difficult, the results did not deviate significantly from the results presented in Fig. 4(a)-(c), in the sense that data points presented in Fig. 4(a)-(c) were not separated from the data points presented in Fig. 4 (d)-(f) and Fig. 4(g)-(i), respectively, by more than two standard deviations. The results of the experiment with both test and comparison stimulus in the high register are presented in Fig. 4(j)-(l). As can be seen in all cases, the rise and the rise-fall had about equal excursion size when lending equal prominence, while the fall had a smaller excursion size than the rise and the rise-fall.

3.2.3 Discussion

The results show that, in order to lend equal prominence, a rise and a rise-fall need to have larger excursion sizes than a fall. In other words, excursion sizes being equal, the fall lends more prominence to a syllable than a rise. This result was obtained for the three accent-lending pitch movements with fixed positions in the syllables. These positions were derived from the IPO text-to-speech system (Collier, 1991) and will be referred to as *standard* positions.

As was shown by the results of the condition in which test and comparison stimulus were presented in different registers, the subjects did not simply compare pitch levels in syllables with pitches on the high declination lines. The larger standard deviations only showed that subjects found the task harder to perform when the stimuli were presented in different registers. As can be seen in Fig. 4(d)-(i), the standard deviations are particularly large for the sessions in which the comparison stimulus was presented in the high register. This appears to be a consequence of the reluctance of some subjects to adjust the excursion size to a relatively high level when it was presented in the high

register. Subjectively, the voice in the high register sounded like a falsetto voice, and giving the pitch movement a large excursion size made it sound extra high pitched, which some subjects apparently tried to avoid.
One of the reasons for the discrepancy between the fall on the one hand and the rise and the rise-fall on the other may be found in their different positions in the syllable. The standard rise and rise-fall started 70 ms before the vowel onset, while the standard fall started 20 ms before the vowel onset. These different positions were obtained from informal listening experiments in the process of obtaining an acceptable intonation in text-to-speech systems. The necessity for a fall to come later than a rise is corroborated by Collier (1970, p. 82). On the basis of results by Van Katwijk and Govaert (1967), he concluded that "a rise is a powerful cue for prominence when it is situated rather early in the syllable, whereas an efficient fall must preferably be located rather late in the syllable." This difference in position might be responsible for the discrepancy.

3.3 EXPERIMENT 2

In this experiment the positions of the rise and the fall were varied in order to find the relation between prominence and position in the syllable. Since varying the temporal position of the rise-fall results in rather unnatural-sounding pitch contours, it was kept fixed. Furthermore, in all cases the rise-fall was either the test stimulus or the comparison stimulus. If all possible combinations of different pitch movements would have been used, the number of trials would have increased to an impossible level. Therefore, one kind of pitch movement was chosen as a reference, the rise-fall in standard position. This rise-fall has a well-defined prominence, and all rises and falls were compared with this pitch movement.

3.3.1 Method

The experimental setup was analogous to that used in the first experiment. The five different temporal positions of the rises and the falls are illustrated in Fig. 5. The position of a pitch movement is given with respect to the vowel onset, as this appears to be the anchor point relative to which an accent-lending pitch movement must be positioned (Collier, 1970; Bruce, 1977). (The vowel onset also appears to be the anchor point in distinguishing between pitch movements characteristic of some Swedish dialects; Bruce, 1983.)

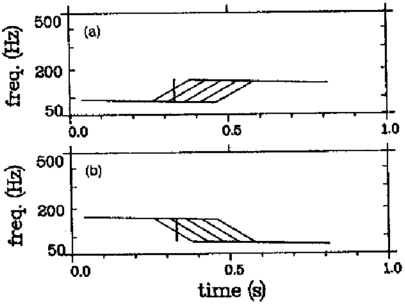


Figure 5. *The five different rises (a) and the five different falls (b) used in experiment 2. The vertical bar indicates the vowel onset.*

The vowel in the accented second syllable lasted 170 ms. The five pitch movements used in this experiment started 70 and 20 ms before the vowel onset, or 30, 80 and 130 ms after the vowel onset. There being two sessions for each position, this resulted in ten sessions for the comparison of the rise with the rise-fall and ten sessions for the comparison of the fall with the rise-fall. All stimuli were presented in the low register. Nine subjects participated in this experiment.

The range of the various positions of the rise and the fall was determined by informal listening. It was chosen so that the pitch movement accentuated the second syllable. As a first impression, the first position of the pitch movement seemed to lend less prominence to the syllable than the second and third positions did, which, however, will not be substantiated by the results. In the last position, again, the prominence seemed to be less.

3.3.2 Results

The results are presented in Fig. 6, in the same way as for the first experiment. Figure 6(a)-(e) present the results for the rise. It appears that, for the two earliest positions, the rise and the rise-fall lent equal prominence if their excursion sizes are equal. For the rises which started later than the vowel onset, however, the prominence became less. Moreover, the variance of the results increased, indicating that subjects found the task more difficult. As shown in Fig. 6(f)-(j), no such effect of position can be observed for the fall. In all five cases, subjects adjusted a smaller excursion size to the fall, indicating that, for equal excursion sizes, the fall lends more prominence to a syllable than a rise-fall. Only in the last position, where the fall started 130 ms after the vowel onset, did the standard deviations tend to be a little larger, indicating that the subjects found the task more difficult to perform.

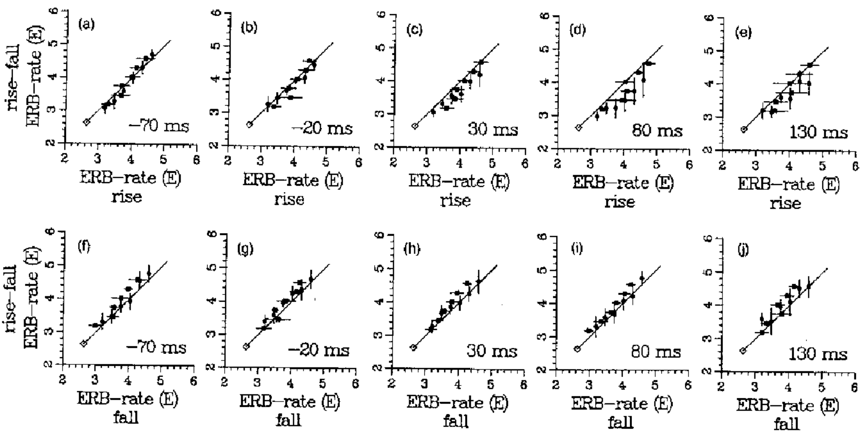


Figure 6. *Results of the nine subjects who participated in experiment 2. In (a), (b), (c), (d), and (e), the standard rise-fall is compared with five rises differing in position in the syllable. In (f), (g), (h), (i), and (j), the standard rise-fall is compared with five falls with different positions in the syllable. The time below the data gives the start of the pitch movement in relation to the vowel onset.*

3.3.3 Discussion

We found that when subjects were asked to adjust the prominence of a fall to the prominence of a rise-fall in standard position, the result was independent of the position of the fall in the accented syllable. In contrast, the timing of the rise was more critical. When the rise started before the vowel onset, the rise and the rise-fall lent equal prominence. When the rise started after the vowel onset, its prominence seems to be reduced and more difficult to compare with the prominence lent by a rise-fall. This might be related to the fact that the two different accent-lending rises in Dutch intonation are distinguished by their timing with respect to the vowel onset. This agrees with the distinction described between an early rise and a late rise in other intonation studies ('t Hart and Collier, 1975; Hill and Reid, 1977). In contrast, the two accent-lending falls in Dutch are predominantly distinguished by their excursion sizes, as described in the Introduction, though position in the syllable plays a role, too ('t Hart *et al.*, 1990).
The results of experiment 2 led us to conclude that the discrepancy between the prominence lent by a rise and a rise-fall on the one hand and a fall on the other could not be attributed to the difference in the timing of the pitch movements used in experiment 1.

3.4 EXPERIMENT 3

In general, the size of prominence-lending pitch movements in Dutch is expressed by the vertical distance between the lower and upper declination lines (e.g., 't Hart *et al.*, 1990). It is then often assumed that, for their standard positions, the amount of prominence lent by pitch movements depends on the size of the pitch change, and that it is independent of the direction of the change. However, it was found in the two previous experiments that falls caused a syllable to be perceived as more prominent than did rises and rise-falls, although they had the same excursion size. The effect was independent of the frequency register in which the test and comparison stimuli were synthesized.
In order to explain the results, a simple model was developed in which the prominence induced by a pitch movement depended linearly on the absolute

difference between the pitch levels at two instants on either side of the pitch movement. These instants might for instance be positioned in the syllabic nucleus of the accented syllable and that of the preceding syllable. These instants will be referred to as *pitch-level estimation points*. The model is illustrated in Fig. 7, which shows a rise and a fall which, according to the original definition being the distance between the declination lines, have equal excursion sizes.

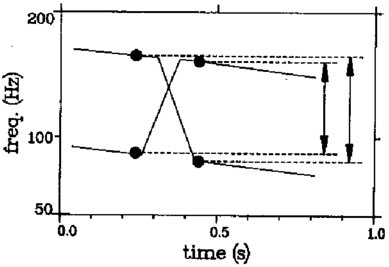


Figure 7. *Diagram showing that for equal excursion sizes the distance between pitch levels in a fall is greater than in a rise. The closed circles indicate pitch levels which might be used for the perceptual determination of prominence. Note that this difference is due to the presence of declination.*

It can be seen that the absolute difference between the pitch-level estimation points is larger when these points are positioned on either side of a fall than when positioned on either side of a rise. It will be clear that this difference S[7] depends on the slope d of the declination line and on the interval T between the pitch-level estimation points. A simple calculation shows this difference to be

$$S = 2dT. \qquad (1)$$

From experiment 1 two quantities in this equation are known; first, the slope of declination d, which was 0.70 E/s and, second, an estimation of S. The latter can be deduced from Fig. 4(c) and 4(l), where it is the average vertical

[7] In the publication the term ΔE was used instead of the term S(hift).

distance between the data points and the diagonal. This average was found to be 0.28 E, from which it can be derived that T is about 0.20 s. This corresponds reasonably well with the distance between the two syllabic nuclei in the stimulus, since the interval between the vowel onsets of the first and the second syllable is 0.22 s. This model predicts that S will increase with increasing declination, and this was the test carried out in experiment 3.

3.4.1 Method

In this experiment rises and falls had standard positions for Dutch as in experiment 1. The same utterance /ma'mama/ as used in experiments 1 and 2 was now synthesized with five different declination slopes: 0.00, 0.33, 0.70, 1.09 and 1.50 E/s, which, in this register, correspond to slopes of 0 and about 2.4, 4.9, 7.3, and 9.7 st/s, respectively. The range of the six excursion sizes of the test stimuli was identical to that in the previous experiments, namely six equidistant values from 0.56 to 1.94 E. In the present experiment, however, the range of the adjustable excursion size of the comparison stimuli, 2.5 E, was divided into eighteen steps, instead of nine, allowing for finer adjustments than in the previous experiments. The small standard deviation found in experiment 1 [Fig. 4(c) and (l)] might be attributed at least partly to the rather large steps, 0.278 E (about 1.67 semitones), between the different comparison stimuli. Thus when subjects adjusted the excursion size in the comparison stimulus one step smaller, the prominence of the accented syllable will have become clearly too small, while adjusting the excursion size one step larger will have made this prominence clearly too large. To get a more appropriate idea of the standard deviation, the steps between the different comparison stimuli were halved.

For each declination, rises were compared with falls and falls were compared with rises, which resulted in ten sessions. Each session comprised six runs, one for each excursion size of the test stimulus. The procedure for presentation of the stimuli and recording of the responses was identical to that used in the preceding experiments. The order of runs within a session was completely randomized. The order of presentation of the various declinations, and within them the order of rises and falls used as test stimuli, was randomized over subjects.

Twelve subjects took part in this .experiment. They were students and research associates of the Institute for Perception Research. All of them reported normal hearing and most of them had experience in auditory perception experiments. Some of them had taken part in the preceding experiments but others had not. In order to determine whether subjects performed consistently, the criterion for consistency described in Hermes and Van Gestel (1991) was applied. As a consequence, the results of three of the twelve subjects were excluded from the analysis.

3.4.2 Results

The results of the nine consistent subjects are presented in Fig. 8. The excursion sizes of the falls are indicated on the horizontal axis, and the excursion sizes of the rises on the vertical axis. The vertical error bars indicate data points obtained when the fall was used as the test stimulus, while the horizontal error bars indicate those for which the rise was the test stimulus. Figure 8(a) presents the data for zero declination.

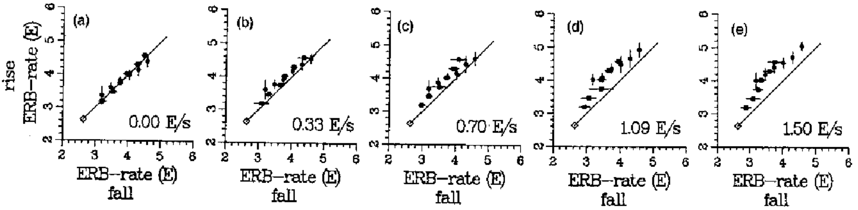


Figure 8. *Results of the nine consistent subjects who participated in experiment 3, in which the slope of declination was varied. This slope is presented below the data in the figures.*

It was found that the data did not differ significantly from the 45-deg line through the origin, indicating that rises and falls of equal excursion sizes lent equal prominence when declination was absent. As can be seen from Fig. 8(b)-(e), a larger declination resulted in an increase of the difference in prominence caused by falls and rises; with equal excursion sizes, falls resulted in greater prominence than rises.

In order to verify whether the results corroborated the predictions according to Eq. (1), the average vertical distances of the data points from the diagonal are shown for each declination in Fig. 9. The straight line represents the predictions based on Eq. (1). For the distance between the pitch-level estimation points T we took the distance between the vowel onsets, namely 0.22 s.

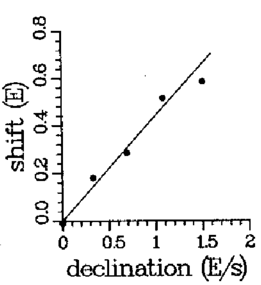


Figure 9. *The data points show the average distance S of the data points presented in Fig. 8 from the diagonal. The line represents Eq. (1), the predicted relation between S and declination. The value chosen for T in this equation was 0.22 s, the distance between the vowel onsets.*

3.4.3 Discussion

In this experiment we tested the effect of declination on the prominence lent to a syllable by a pitch movement. It was assumed that this prominence depends linearly on a difference in pitch level at some point before the pitch movement and at some point after the pitch movement. For the interval between these pitch-level estimation points we chose the interval between the vowel onsets. It was found that the model predicted the results of the experiment quite accurately.

The experiment with a declination slope of 0.70 E/s is identical to the experiment with results shown in Fig. 4(c), except that the step size between the various comparison stimuli was halved. It will be noted that the standard deviation is now larger. This indicates that the small standard deviation shown in Fig. 4(c) can indeed be attributed to the large step size used there.

From Eq. (1), it follows that S, the distance between the data points and the diagonal in the figures, should be independent of the excursion size of the test stimulus. Especially for the larger declination slopes, however, it can be seen in Fig. 8 that the data points are not all the same distance from the diagonal. More specifically, the adjustments for the smaller excursion sizes of the test stimulus are systematically larger, while those for the greater excursion sizes are smaller. This finding probably represents a systematic trend towards the average of the results. The fact that this trend was larger for the larger declination slopes, indicates the task is more difficult then. Since the upward bias in the adjustments for the smaller test stimuli will be compensated by the downward bias in the adjustments for the larger test stimuli, this will not have influenced the average distance from the diagonal, which was the quantity in which we were interested.

3.5 GENERAL DISCUSSION

First, we will discuss two findings which were quite unexpected. The first is that the timing of the fall did not affect the prominence of the accented syllable. The second is that, excursion sizes being equal, a fall lends more prominence to a syllable than a rise and a rise-fall of equal excursion size. Then we will discuss the implications of the model we suggested for the perception of prominence in speech intonation and the role of pitch movements in this respect.

3.5.1 Timing

That the prominence of a fall compared with that of a standard rise-fall was independent of the position of the fall was unexpected, because informal listening gave the impression that a fall starting 70 ms before the vowel onset lent less prominence than a fall starting 20 ms before or 30 ms after the vowel onset. The easy way out is to say that varying the temporal position of the pitch movement affects only the conspicuousness of the pitch movement but not the prominence of the syllable it accentuates. For the rise, for which an effect of timing was found, the results imply that its timing does not actually affect the prominence of the accented syllable, but only determines the type of pitch movement with which the syllable is accentuated. In other words, the early rise is phonologically, and hence perceptually different from the late rise. The advantage of this explanation is that it removes timing from the list of factors that determine the prominence of a syllable. In that sense it reduces the number of factors we must investigate in order to unravel the physical background underlying prominence perception in intonation.

The actual situation will be much more complicated, however. In a different experimental setup, Rump (1992) found evidence that subjects perceive a fall ending before the vowel onset as less prominent than a later fall. He varied the timing of a pitch movement and the duration of the vowel in the accented syllable. The discrepancy between this result and ours indicates that the physical attribute underlying prominence perception is multidimensional and that one dimension cannot simply be traded off for another.

A possible definition of prominence may run ''the perceptual conspicuousness of a syllable'', in which the conspicuousness is then determined by various dimensions. A similar situation might be found in reading research, e.g., when subjects are asked to equate the conspicuousness of a letter in a text with varying luminance and contrast. If subjects are asked to adjust the luminance of one letter to that of another until they have the same conspicuousness, they will concentrate on the brightness of the letters, and disregard perceived contrast. If, on the other hand, subjects are asked to adjust the physical contrast of one letter to that of another until the same contrast is perceived, they will concentrate on the perceived contrast of the letters and disregard brightness. In the same way, pitch being the adjustable dimension, subjects in the present experiment 2 may have paid attention only to prominence as induced by the excursions of the pitch movements and may have disregarded their timing. The results reported in this paper consequently do not necessarily demonstrate that the timing of pitch movements is irrelevant to the prominence of the syllable. They do show that subjects can compare the prominence of some particular accent-lending pitch movements. In addition, the subjects appear to be very consistent in paying attention to something related to excursion size, and they apparently stick to this dimension when comparing the stimuli. In experiments in which subjects have a different task, e.g., to indicate in which of two utterances the accented syllable has more prominence, they may very well pay attention to several additional dimensions such as the timing of the pitch movement or the duration and amplitude of the vowel. In such cases the outcome of the experiments will very much depend on the dimension the subject directs his or her attention to. Comparison of different experimental paradigms may clarify this issue.

3.5.2 Excursion sizes

We found that, while lending equal prominence, the excursion size of the fall was smaller than that of an early rise or a rise-fall. These results might seem to conflict with results obtained by Van Katwijk and Govaert (1967), who synthesized four different vowels interrupted by noise, resulting in something which sounded like /χiχoχεχyχ/. The pitch contour of this utterance contained either a rise or a fall, superimposed on a declination line starting at 160 and ending at 110 Hz. They systematically varied the position of this pitch movement in the syllable and asked subjects to indicate which syllable carried an accent. For the rise, the result was as expected: The syllable with the pitch movement was almost always perceived as accentuated. For the fall, however, the result was less clear. In a considerable number of cases, subjects chose the first syllable as the accentuated one, though the fall was positioned on a later syllable. Van Katwijk and Govaert concluded that the fall lent less prominence to a syllable than a rise. However, they had provided their utterance with a very strong declination, namely about 1.2 E/s. The start of their utterance consequently had a rather high pitch. Furthermore, the utterance began with an unvoiced consonant. In combination with the succeeding fall, this may have given the impression of a virtual rise at the start of the first syllable. This virtual rise would then have accented the first syllable, as indicated by the subjects. In our experiments, declination was 0.70 E/s and the utterance started with a voiced sonorant consonant. In addition, subjects were explicitly asked to direct their attention to the prominence of the second syllable. This precluded the perception of a virtual rise preceding the first syllable. So, with the mentioned re-interpretation of Van Katwijk and Govaert's results, there is no disagreement between their results and ours.

3.5.3 A model for prominence caused by pitch movements

Relatively little is known about the intonational factors determining the perception of prominence in speech. House (1990) developed a model according to which pitch movements will be coded by the listeners either as movements or as levels. For a pitch movement to be coded as a movement three conditions should be fulfilled. First, a large part of the movement should take place in an interval of relative spectral stability, i.e., the syllabic nucleus. Second, the beginning of the pitch movement should start 30 to 50 ms after the vowel onset. Third, the vowel should last at least 100 ms. If these conditions are not fulfilled, the pitch movements will be coded as a change in pitch level. According to this model, the early pitch movements in the present experiments will be coded as levels rather than as movements since they start before the vowel onset. The late pitch movements, in any case those starting 80 and 130 after the vowel onset, will be coded as movements, since the accented vowel lasted 170 ms. This may explain the different results found for the early rise and the late rise. It is unclear, however, why such a discrepancy was not found for early and late falls.

In the model presented and verified in experiment 3 it is assumed that the prominence of a syllable is derived from relations between two successive pitch levels, indicated by the filled circles in Fig. 7, on either side of the pitch movement. Owing to declination, the pitch interval between these two pitch-level estimation points is larger for the fall than for the rise. This may explain why, in the presence of declination, a fall lends more prominence to a syllable than a rise. Moreover, this may explain why a larger slope of declination results in a more prominent fall, and a less prominent rise. A direct consequence of this is that, at the relatively short time scale of these accent-lending pitch movements, the perceptual process presumed to underlie the perception of prominence does *not* compensate for declination. It is hard to say how this can be brought in line with the compensation for declination found for pitch levels in successive accented syllables within one utterance (Pierrehumbert, 1979). Also, the findings by Terken (1991) are not in quantitative agreement with the proposed model. He found that, when subjects were asked to adjust the second of the two successive rise-falls within one utterance so as to give it the same prominence as the first, they did not give it an equal excursion size on an ERB-rate scale. This shows that the perception of the prominence of an accented syllable in an utterance may depend on the prominence of previous or next syllables and on long-term declination.

It follows from Eq. (1) that S is independent of the excursion size of the rise and the fall. This explains why the data points displayed in Fig. 4 run parallel to the diagonal. This has a strange consequence when the data points are extrapolated to excursion size zero: A fall of excursion size zero will be adjusted to a rise with a positive excursion size. This implies that both still have some prominence though the fall has excursion size zero and that the rise might lend no prominence though its excursion size is nonzero. This, then, shows that prominence may not be determined only by excursion size. In addition to the primary cues, duration and intensity, declination in itself can increase the prominence of a syllable. In this situation, a clear pitch accent is perceived on the second syllable even in the absence of any decrease in pitch other than is due to declination. This agrees with informal listening to stimuli of this kind. A prerequisite, however, is that the accent on the second syllable is obvious even in the absence of a clear pitch movement. The utterance /ma'mama/ used in these experiments was spoken with a clear accent on the second syllable, so that amplitude and probably duration also play an important role in the accentuation of the second syllable.

It is not yet clear how long the time interval should be between the two pitch-level estimation points which the listeners compare in order to determine the difference in pitch level. From the results of experiments 1 and 2 it can be deduced that this difference is derived from the pitch levels in the accented syllable and the preceding syllable, and not from the pitch levels in the accented syllable and the following syllable. This follows from the similarity between the rise and the rise-fall. If the level difference between the accented syllable and the following syllable played a major role, the prominence lent by a rise-fall would be similar to that lent by a fall. In the model presented here it is implicitly assumed that the pitch-level estimation points are in the syllabic nuclei. There may, however, very well be a fixed distance which accidentally corresponds quite well with the distance between the nuclei of succeeding syllables in normal speech. In the stimuli used in these experiments the time interval between succeeding nuclei was 0.22 s. In fluent speech it has the same order of magnitude. Future experiments will have to clarify whether the distance between the pitch-level estimation points is fixed, or whether it varies with the actually realised distance between the syllabic nuclei of fluent speech.

3.6 CONCLUSIONS

Experiment 1 shows that subjects are quite capable of comparing the prominence of syllables accented by three different accent-lending pitch movements in standard position: A rise starting before the vowel onset, a rise-fall and a fall. In addition, excursion sizes being equal, the fall lends more prominence to a syllable than a rise or a rise-fall. Experiment 2 ruled out the possibility that the different temporal position of the pitch movements underlies this difference between the fall on the one hand and the rise and the rise-fall on the other. In the case of the fall, varying the temporal position does not influence the perceived prominence relative to a rise-fall in standard position. As to the rise, subjects have difficulty in comparing the prominence induced by a late rise with the prominence of a rise-fall in standard position. This might be a consequence of the phonological difference in Dutch between an early rise, a rise-fall and a fall on the one hand, and a late rise on the other. Within and across the phonological categories consisting of the accent-lending pitch movements, the early rise, the rise-fall and the fall, prominence is a well-defined concept, in the sense that subjects are well able to compare the prominences of the syllables accented with pitch movements from one of these categories. The late rise apparently belongs to a different category, so that the prominence it lends cannot be compared with the prominence lent by a pitch movement from the other category.

The reported experiments were carried out for an utterance with one accented syllable. There is no reason to believe that the conclusions should not be valid for utterances with more than one accented syllable. As mentioned earlier, a succession of an early rise and a fall on successive accented syllables can be replaced by two rise-falls without affecting the prominence of the accented syllables. Quantitatively, however, we must take into account that, at the relatively short time scale of the utterances described in this paper, we do not compensate for declination perceptually.

The prominence of an accented syllable depends on at least four dimensions: the excursion size of a pitch movement, the timing of a pitch movement, the duration of the vowel and the intensity of the syllable. In experiment 2 it was found that the contribution of pitch cannot simply be traded off for the contribution of timing. A good estimation of the contribution of a pitch movement to the prominence of an accented syllable is the difference in the number of ERBs between the average pitch of the syllabic nucleus of the accented syllable and the average pitch of the syllabic nucleus of the preceding syllable.

4 PROMINENCE LENT BY RISING AND FALLING PITCH MOVEMENTS: TESTING TWO MODELS[8]

ABSTRACT

Two experiments are reported in which a pitch-level-difference (PLD) model for prominence perception (chapter 3; Hermes and Rump, 1994) is subjected to further tests. The model holds that the contribution of pitch to the perceived degree of prominence is proportional to the difference in pitch level between the vocalic nuclei of the accented and the previous syllable. In experiment 1, the influence of stretching and compressing the utterance in time was assessed. It was found that the predictions made by the model were not fully supported by the data. An alternative model was developed according to which pitch movements resynthesized in the same register lend equal prominence when pitch levels on the upper declination lines in the stimuli are equal. These two models gave different predictions when the lower declination lines are different. This was tested in experiment 2. The results which are more or less between the predictions by the two models suggest that low pitch levels play a smaller role in prominence perception than high pitch levels do.

4.1 INTRODUCTION

Prominence, defined as the conspicuousness of a syllable, plays an important role in intonation. Various studies have been reported relating the relative prominence of pitch-accented syllables to the succession of pitch peaks (e.g. Bruce, 1982). A typical conclusion of one such study is, e.g.:

"After a focal accent, the downstepping of successive non-focal accents is a characteristic

[8] Published as H.H. Rump and Dik J. Hermes: "Prominence lent by rising and falling pitch movements: testing two models," Journal of the Acoustical Society of America **100** (1996), 1122-1131. The results of experiment 2 were also published as H.H. Rump and Dik J. Hermes: "Comparison between two models for prominence perception," in the IPO Annual Progress Report **29** (1994), pp. 29-35, and they were presented at the 128th meeting of the Acoustical Society of America, Austin, Texas, as H.H. Rump and Dik J. Hermes: "Some control experiments on a model for prominence perception," Journal of the Acoustical Society of America **96** (1994), 3349.

> pitch pattern. This downstepping seems to be the expression of equal prominence of successive post-focal accents within the phrase. In a pre-focal position, however, up to the focal accent of a phrase (or of a whole utterance) there is typically no downstepping, but instead only a very gentle declination (if any) for successive non-focal accents."
>
> (Bruce and Touati, 1992, pp. 455-56).

Careful studies to relate equal prominence of accented syllables within one utterance to simple acoustic features of the pitch contour have not lead to a simple and plausible model (Pierrehumbert, 1979; Rietveld and Gussenhoven, 1985; Terken, 1991, 1993b, 1994; Ladd, 1993; Repp *et al.*, 1993). This may be due to the implicit focus structure of the utterance coupled with the absence or presence of downstep, which may have been confounding factors. In addition, the proposed models deal with the relative heights of the rise-falls which are associated with the accented syllables, leaving other possible accent-lending pitch movements like isolated rises and falls aside.

Only very few studies (Hermes and Van Gestel, 1991; Hermes and Rump, 1994) have been reported dealing with the relative prominence of accented syllables in two successive utterances. In these studies, the pitch accents were always on the second syllable of the three-syllabic utterance /ma'mama/. Hermes and Rump (1994) asked subjects to adjust the excursion size of the pitch movement in the second utterance until the prominence in that utterance was equal to the prominence of the pitch-accented syllable in the first utterance. The utterances were resynthesized in the same or in different registers. The pitch movements in the two utterances, however, were always different. Each could be a rise, a fall, or a rise-fall. One of the main findings was that, lending equal prominence, a fall had a smaller excursion size than a rise or a rise-fall. Varying the temporal position of the rise only lead to different results when the rise started later than the vowel onset of the accented syllable. Varying the temporal position of the fall did not affect the results very much. In the description of Dutch intonation by 't Hart *et al.* (1990) an accent-lending rise can be either early (Rise 1) or late (Rise 3), in autosegmental phonology corresponding with L+H* and L*+H, respectively. In the description by 't Hart *et al.* there is only one full-sized fall, corresponding with H+L* in autosegmental phonology.

The unexpected finding in the experiments by Hermes and Rump (1994) was that excursion sizes being equal, the rise and the rise-fall lent less prominence to an accented syllable than the fall. This did not depend on whether the two utterances were in the same register or in different registers. Based on these considerations, Hermes and Rump proposed a pitch-level difference (PLD) model in which the prominence lent by an accent-lending[9] pitch movement is proportional to the difference between the pitch level[10] in the vocalic nucleus[11] of the accented syllable and that of the preceding syllable. In other words, the height of the pitch level in the accented syllable is judged relative to the height of the pitch level in the preceding unaccented syllable. The difference between the two levels is the PLD. The PLD model states that two pitch movements lend equal prominence when the PLDs are equal.

The PLD model is illustrated in Fig. 1, in which two accent-lending pitch movements, a rise and a fall lending equal prominence, are shown on top of each other. It explains the (experimentally observed) difference in excursion size needed to match the prominence lent by the two pitch movements. 'Excursion size' is defined, like in 't Hart *et al.* (1990), as the distance

[9] Whether a pitch movement is accent-lending is determined by its position in the syllable and by whether it is abrupt or gradual, although this is largely language specific (for Dutch see 't Hart *et al.*, 1990). In pitch-accent languages like Dutch or English, a syllable is called 'pitch-accented' if its prominence relative to its neighbours is mainly due to the presence of an accent-lending pitch movement, which is mostly accompanied by a longer syllable duration, higher amplitude, and/or better vowel quality. Changes in pitch are normally perceptually more salient than changes in one of the other accent-lending factors. Speakers may vary the degree of prominence of accented syllables for linguistic (e.g. focus) or paralinguistic (e.g. emphasis, emotions) reasons. Accent as used in the present paper should not be confused with lexical stress. A stressed syllable can or cannot be marked by an accent-lending pitch movement. Beckman (1986) presents evidence that longer duration, higher amplitude and/or better vowel quality are acoustically correlated with stress, at least in English (and in Dutch). As Pierrehumbert (1980; also Beckman and Pierrehumbert, 1986) makes clear, these properties also correlate with the presence of a pitch accent, but only so because the pitch accent lines up with a stressed syllable.

[10] According to House (1990), pitch movements are only perceived as such if they start after the vowel onset, if they have a minimum duration of about 100 ms, and if they take place in spectrally stable signals like vowels. Otherwise, rises are 'perceptually recoded' to high pitch levels and falls to low pitch levels.

[11] The vocalic nucleus is defined as the area of maximum sonority of the syllable. It may also be called the syllabic nucleus. Note, however, that the vocalic nucleus is not the same as the P-center of the syllable (e.g. Pompino-Marschall, 1990). The former is necessarily in the vocalic part of a syllable, whereas the P-center is defined as the perceptual moment of occurrence of the syllable, which may be even before the onset of the vocalic part of the syllable.

between the lower declination line, or baseline, and the upper declination line, or topline, measured perpendicularly to the time axis. Here, in contrast to 't Hart *et al.* who measured excursion sizes in semitones, the distance is measured in numbers of equivalent rectangular bandwidths (ERBs; unit E) (Patterson, 1976; Hermes and Van Gestel, 1991). Declination lines run parallel in the ERB-rate-frequency domain.

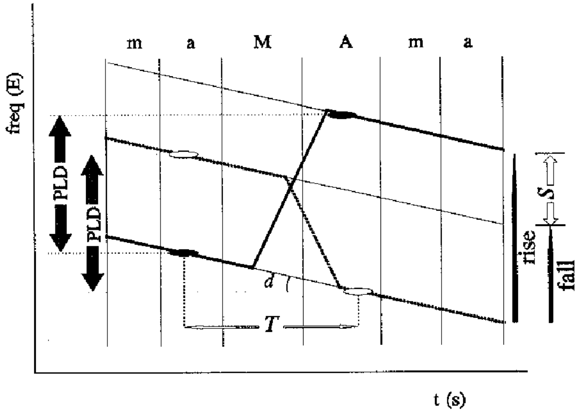


Figure 1. *Illustration of the PLD model. The pitch contours of a rise and a fall have been plotted one on top of the other. The vertical lines indicate the segmental boundaries (time axis almost identical to original utterance). It can be seen that the excursion sizes (black arrows on the right) of the rise and the fall differ by the amount S (shift, open arrow on the right) when their pitch-level differences (PLDs, arrows on the left) are equal. The relevant pitch levels are indicated by ellipses. Symbol T indicates the time interval between the pitch levels, symbol d indicates the rate of declination.*

Pitch levels that are assumed to be relevant for the determination of PLDs are indicated by ellipses. The pitch averaged across the vocalic nuclei, which are defined as areas of maximum sonority, were believed to provide the relevant pitch levels. The black ellipses represent the relevant pitch levels for the rise, the open ellipses represent the relevant pitch levels for the fall. The arrows on the left-hand side indicate the PLDs for the rise and the fall. The length of the arrows may hence be seen to represent the contribution of pitch to the prominence lent by a rise or a fall.

From the traditional view of prominence perception (e.g. Cohen *et al.*, 1982), it would follow that a rise and a fall lend equal prominence when their excursion sizes are equal, i.e. that listeners take declination into consideration. In the PLD model, however, no such compensation for short-term declination, i.e. declination over the time span of two syllables within one utterance, is assumed. In Fig. 1 the black arrows on the right-hand side indicate the excursion sizes of the rise and the fall lending equal prominence. The difference in excursion size (open arrow) is called the ''shift'' *S*, which can be calculated to equal twice the rate of declination *d* times the time interval *T* between the relevant pitch levels, so that

$$S = 2dT \tag{1}$$

Shifts are measured in ERB-rate units E since *d* is measured in E/s and *T* is measured in s. In this equation, the declination d is assumed to be constant over the utterance, or at least over the syllable preceding the accented syllable and the accented syllable itself.

Hermes and Rump have shown that the predictions derived from Eq. 1 were quite accurate for different declination rates, *d*. They measured *S* as a function of *d* and so they could calculate *T*. It was found that the measured shifts were indeed linearly related to the rate of declination. The value of *T* appeared to be about 220 ms (*S/2d*) which corresponded more or less precisely with the distance between the vowel onsets of the syllabic nuclei of the accented syllable and the preceding syllable. Since the shift depends on both the rate of declination and the time interval between relevant pitch levels (Eq. 1), the PLD model predicts that, if *d* is kept constant and is not zero, a change in *T* will also affect the PLDs. For example, an increase in *T*, achieved by slowing down the speech rate, will result in a larger shift, because, at the same time, the PLD for the fall increases, while the PLD for the rise decreases. It is not known, however, whether the relevant pitch levels are indeed anchored to the vocalic nuclei. *T* may also be fixed to some value of about 220 ms. According to the autosegmental analysis of bitonal pitch accents, a fixed value for *T* may be assumed. Pierrehumbert (1980, p. 77), for example, found for the L*+H- accent that the L* tone is directly associated with the accented syllable, while the position of H- is fixed at about 200 ms after the starred tone, more or less independently of the syllabic structure. The situation in Hermes and Rump's experiments was different from that in Pierrehumbert's. Pierrehumbert obtained the value for a bitonal accent in which the first and not the second tone was associated with the stressed syllable, whereas in Hermes and Rump, it was the second tone which was starred, i.e. H+L* = falling and L+H* = rising. However, it may be that a leading unstarred tone also occurs at a fixed interval from the starred tone, like a trailing unstarred tone does. If T is indeed fixed, changing the speech rate might not result in a change in the shift. The aim of the first experiment was to test this.

The exact distance between vocalic nuclei is difficult to determine. As the time interval T turns out to be about the same as the time interval between the vowel onsets of the accented and the previous syllable V, this time interval V between subsequent vowel onsets is used for predicting the shift:

$$S = 2dV \tag{2}$$

The hypothesis is then that the inter-vowel-onset interval V and the inter-pitch-level interval T are identical, i.e. $T = V$. If T is indeed anchored to the vowel onsets, shifts are predicted to be equal to $S = 2dT = 2dV$, i.e. S will increase if V increases, and S will decrease if V decreases, the non-zero declination being kept constant. Alternatively, if the T is a stable interval, shifts are predicted to be independent of V, and hence to be equal to $S = 2dT$ (Eq. 1), and not to $S = 2dV$ (Eq. 2).

4.2 EXPERIMENT 1

4.2.1 Method

In Hermes and Rump it was shown that listeners were well able to adjust prominences to make them equal if they were associated with pitch movements which were resynthesized in equal or different pitch registers. The measured shifts turned out to be almost the same, regardless of the register. Only the standard deviations of the shifts were larger for different registers, 0.2 E, than for equal registers, <0.1 E. It was concluded that, in general, the same averages would be obtained for different registers as for equal registers, and that only standard deviations would differ. Therefore, in the current experiment, the stimuli were resynthesized in the same register, since this would lead to smaller standard deviations. This was important because changing the inter-vowel-onset interval, *V*, was expected to result in relatively small differences between the measured shifts under the different conditions. The predicted shifts under the different stimulus conditions were as follows: The declination rate, *d*, was fixed at 1.0 E/s and the inter-vowel-onset intervals, *V*, were 150, 220, and 350 ms. The ratio between the short *V* and the long *V* was therefore greater than 1:2. According to Eq. 2, the predicted shift, *S*, was 0.30 E for the short stimuli, 0.44 E for the original stimulus duration, and 0.70 E for the long stimuli. Shifts under the different stimulus-duration conditions were therefore expected to differ by about 0.4 E at most. Large standard deviations might obscure these small differences.

4.2.1.1 Stimuli. The stimuli consisted of modified versions of the carrier (nonsense) utterance /ma'mama/, spoken by a male speaker, which was the same utterance as that used in Hermes and Rump. Manipulations of duration and pitch were performed using the pitch-synchronous overlap-and-add (PSOLA) method developed by Hamon *et al.* (1989), resulting in natural-sounding speech. Vowel onsets in the original utterance were determined perceptually. In order to make the short and long stimuli sound optimally natural, the utterance was shortened and lengthened non-linearly, as is illustrated in Fig. 2. The resulting values of *V* were 150 and 350 ms for the short and the long stimuli, respectively. In the stimuli with the original syllable durations *V* was 220 ms. The extra lengthening of the /m/ in the second syllable of the long stimuli in particular seemed to be crucial for the second syllable still to be perceived as accented. In addition, this extra lengthening helped to increase *V* without making the stimuli sound unnatural.

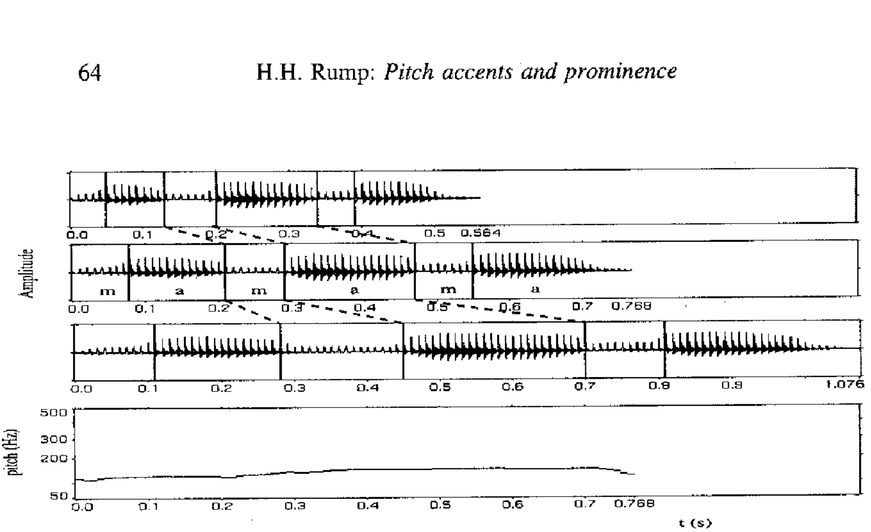


Figure 2. *The upper three panels show the waveforms for the three stimulus durations, shortened, original, and lengthened, respectively, used in experiment 1. The vertical lines indicate phoneme boundaries. The dashed lines connect the corresponding phoneme boundaries of the accented syllables in the three duration versions. The lower panel shows the pitch contour of the original utterance.*

An accent-lending rise and fall were effected between the baseline and the topline which ran parallel in the ERB-rate domain. The pitch movements had durations of 120 ms, independent of the speech rates. The rise started 70 ms before the vowel onset of the second, accented syllable, the fall started 20 ms before the vowel onset of the accented syllable (the standard values for rise 1 and fall A in the IPO text-to-speech synthesis system, see 't Hart *et al.*, 1990; Terken, 1993a). The end frequency of the baseline was fixed at 75 Hz (2.64 E). The rate of declination was also the same for all stimulus durations, i.e. 1.0 E/s. This was somewhat higher than the standard declination rate of 0.7 E/s. Since declination is a factor in the PLD model, a relatively fast rate of declination would make the differences between predicted shifts as large as possible while keeping the sound of the stimuli natural. Since the rate of declination and the end frequency of the baselines were kept constant during the experiment, the starting frequencies of the baselines for the three stimulus durations were different, i.e. 93 Hz (3.18 E), 101 Hz (3.39 E), and 111 Hz (3.69 E) for the short, original, and long stimuli, respectively.

The stimuli were presented in pairs. The first stimulus will be referred to as *test* stimulus, the second one as *comparison* stimulus. The test stimulus had four excursion sizes: 0.83, 1.11, 1.38, and 1.67 E. The range of adjustable comparison stimuli, from 0 to 2.5 E (about 94 Hz or 14 semitones), was divided into 18 equidistant steps of 0.14 E.

4.2.1.2 Procedure. The test and comparison stimuli always had identical durations. Between the stimuli in a pair there was a silent interval of about 600 ms. The listeners matched the adjustable excursion size of the pitch movement in the comparison stimulus until its prominence matched the prominence lent by the opposite type of pitch movement in the test stimulus. Within one trial, the excursion size of the latter was fixed while the comparison stimulus had a variable excursion size. This means that after listening to a stimulus pair, the listeners could increase or decrease the excursion size of the comparison stimulus by pressing keys on a computer key board. They could repeat this until they found that the prominences of the accented syllable in the comparison stimulus and that in the test stimulus were equal. Their key strokes were recorded automatically. When they were satisfied by the finally adjusted excursion, they pressed the return key and the next stimulus pair was presented. A typical number of key strokes within a trial was nine.

For each excursion size of the test stimulus, the comparison stimulus was adjusted twice, starting from either end point of its excursion size continuum. This was done in order to counter the effect of hysteresis, a methodological bias due to which excursion sizes are made larger when the starting point is high and smaller when the starting point is low[12]. In other words, for both rising and falling comparison stimuli, larger adjustments in excursion are made when starting at the high than at the low end of the continuum. Another way of undoing the effect of hysteresis is to take the means of adjustments of the

[12] Hysteresis in the classical sense is defined as the limping-behind of one factor relative to another, causally related one. This would imply that the adjusted excursion size would be larger if the starting point is low and smaller if it is high. This is not what we found. Here we use the term hysteresis to indicate that the excursion size at the first presentation of the stimuli seems to influence the total range of ''acceptable'' excursion sizes. As a result this range is relatively small when the starting point is low and relatively large when it is high. This might explain why we found that adjusted excursions are smaller when the starting point is low than when it is high.

rise and the fall, as was done in Hermes and Rump. Taking two starting points for each adjustment, however, leads to better estimates of the adjusted excursion sizes for both rises and falls. The total number of trials was 48 (3 stimulus durations x 2 pitch movements in the test stimulus x 4 excursion sizes x 2 starting points). The order of presentation was randomized.

Nine subjects participated in the experiment. They were instructed to make the prominences ('pep' or 'schwung') of the two stimuli equal. All of them had taken part in earlier prominence-adjustment experiments. The consistency of their scores was determined using the criterion that the correlation between the scores under a given stimulus-duration condition had to reach the level of 0.75 (Hermes and Van Gestel, 1991). Data were excluded if subjects did not reach the level of consistency under at least two of the three conditions.

4.2.2 Results and discussion

One of the subjects mentioned that he had adjusted equal musical pitch intervals in the individual stimuli, while not paying much attention to prominence. His data were therefore omitted from any further analysis. The data of two other subjects did not reach the required level of consistency, and they were therefore also discarded.

The analysis of the results of the six remaining subjects showed that the effect of hysteresis was significant ($F_{(1,3)} = 42.5$, $p < 0.008$). The adjusted excursion sizes were about 0.14 E larger when adjustments started at the high end of the continuum than when adjustments started at the low end of the continuum. The magnitude of the effect was equal to about one step in the excursion-size continuum. Since we wanted to counter the effect of hysteresis, the results were pooled across starting points.

Figure 3 shows the adjusted excursion sizes averaged across the six subjects. Panel (a) shows the results for the short stimuli, panel (b) for the stimuli having the original duration, and panel (c) for the long stimuli. The horizontal axes give the excursion sizes of the rise, the vertical axes give the excursion sizes of the fall. The mean adjusted excursion sizes of rises, for which the test stimulus contained a fall, are indicated by open circles and horizontal standard-deviation bars, those of falls, for which the test stimulus contained a rise, are indicated by black circles and vertical standard-deviation bars. The standard-deviation bars indicate the combined effect of intra-subject variance and inter-subject variance.

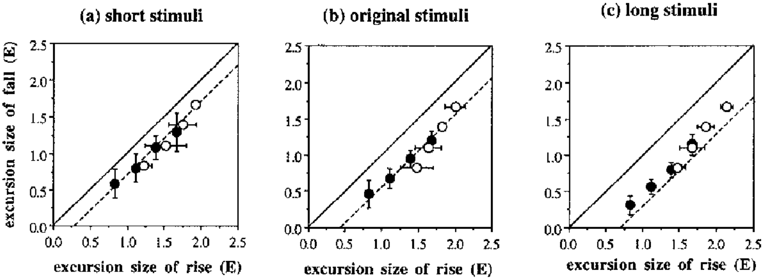


Figure 3. *Results of the six consistent subjects who participated in experiment 1. Panels (a) to (c) show the results for the short, original, and long stimulus durations, respectively. The black circles and vertical standard-deviation bars show the results for the fall, the test stimulus being on the horizontal axis, the open circles and horizontal standard-deviation bars show the results for the rise, the test stimulus being on the vertical axis. The diagonal indicates equal excursion sizes, the dashed line indicates the values predicted by the PLD model.*

As in the experiments reported by Hermes and Rump, the results displayed a trend towards the average in the stimulus range. This bias often occurs in adjustment experiments. It can be very strong in difficult tasks. In the extreme, listeners will select a random stimulus from the whole stimulus range. The average of these adjustments would be at the average of the range. Another explanation for this trend may be that subjects, after having listened to many stimuli, have built up a kind of ''standard'' stimulus of average excursion size. At low attention levels or in difficult tasks, the adjustments of the subject would be drawn into the direction of this standard stimulus. Note that, in the present experiment, by the way the data are presented, this trend towards the average is in the vertical direction for the rises and in the horizontal direction for the falls. As a result, the regression lines for rises and falls do not coincide. Therefore, in order to correct for this trend towards the average, the data for rises and falls are combined, resulting in a better estimate of the shift.

The solid line in each of the panels of Fig. 3 indicates equal excursion sizes of the pitch movements. The distance between the solid and the dashed line, measured perpendicularly to the horizontal axis, is the predicted shift: $S = 2dV$ (Eq. 2). The dashed line in each of the panels therefore indicates the positions of the measurement points that would be predicted if the time interval between relevant pitch levels T and the time interval between subsequent vowel onsets V were to be identical. It can be seen that the majority of the measurement points lie along the dashed lines. A repeated-measures ANOVA showed that the effect of V on S was significant ($F_{(2,10)} = 25.3$, $p < 0.001$). As predicted by the PLD model, a decrease in V caused a decrease in the shift, and an increase in V caused an increase in the shift.

The shifts for the individual subjects are shown in Table I. A longer V resulted in a larger shift in the case of almost every individual subject. In the case of Subject 4, the effect of V on the adjusted excursion sizes was marginal with respect to the shortening of the stimuli, and indeed there was no effect with respect to the lengthening of the stimuli and his shifts were consequently more or less constant. Mean shifts and their standard deviations were calculated from the distances between the individual data points and the diagonal, measured perpendicularly to the horizontal axis, in each of the panels in Fig. 3. The predicted shifts were based on Eq. 2.

The mean shift calculated for the short stimuli turned out to be somewhat larger than predicted. The mean shift calculated for the long stimuli were considerably smaller than predicted (see also panel c in Fig. 3). The values of T, the time interval between relevant pitch levels, which could be derived from the measured shifts ($T=S/2d$), were about 170, 225, and 270 ms for the short, normal, and long stimuli, respectively.

Table I. *Individual shifts (S) in experiment 1, the shifts predicted by the PLD model (Eq.2), and the mean observed shifts with their standard deviations. Measures in number of ERBs.*

	stimulus duration		
	short	original	long
Subj.			
1	0.295	0.399	0.530
2	0.365	0.547	0.590
3	0.156	0.278	0.399
4	0.503	0.521	0.521
5	0.356	0.530	0.634
6	0.330	0.443	0.564
pred. (Eq. 2)	0.300	0.440	0.700
obs.	0.334	0.453	0.540
sd	0.056	0.093	0.055

The corresponding values of *V* were 150, 220, 350 ms. *T* was therefore 20 ms longer than *V* in the short stimuli, whereas it was 80 ms shorter than *V* in the long stimuli. (In the case of the original stimuli, the finding of Hermes and Rump that *T* and *V* were about equal was replicated.) The decrease in *T* was therefore only 15 ms smaller than the decrease predicted by Eq. 2 (55 instead of 70 ms), but the increase in *T* was 85 ms smaller than the predicted increase (45 instead of 130 ms). This means that the predictions derived from the PLD model were not fully supported if it is assumed that the distance between the relevant pitch levels stretches and shrinks with the distance between the vowel onsets. Alternatively, it may be that the shrinking of *T* indeed takes place, but that the stretching of *T* is somewhat less than that of *V*. In other words, *T* increases with an increase of *V*, but the change is somewhat smaller.

4.2.2.1 A different listening strategy: pitch levels of the same heights. So, it may still be possible to explain the results of the experiment within the framework of the PLD model. Based on the comments by some of the subjects as well as for theoretical reasons, however, an alternative strategy will be presented to account for the discrepancy between predicted and observed shifts. Given that the task of adjusting prominence was a rather difficult one, some subjects mentioned that they may have been following the strategy of adjusting the high pitch levels in a stimulus pair so as to be equal. In the following section, we will present a quantitative analysis of the shifts predicted by the strategy of matching high pitch levels, which may give a better explanation of the present findings. In experiment 2, the new predictions will be contrasted with those based on the PLD model.

If we compare the predicted outcome of this strategy to the predictions of the PLD model, the following occurs. According to the PLD model, listeners adjust the PLD of the comparison stimulus so as to make it match the PLD of the test stimulus.

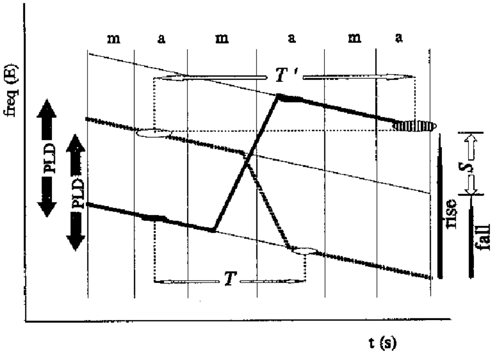


Figure 4. *As Fig. 1. It can be seen that the height of the high pitch level in the unaccented syllable before the fall (open ellipsis on the left) and the high pitch level after the rise (hatched ellipses on the right) are about equal when the PLDs (arrows on the left) are equal. T indicates the time interval between the pitch levels in subsequent syllables, T' indicates the time interval between the high pitch levels in the first syllable, i.e. before the fall, and in the third syllable, i.e. after the rise. As in Fig. 1, the excursion sizes (black arrows on the right) of the rise and the fall then differ by the amount S (shift, open arrow on the right).*

This is illustrated in Fig. 4, in which the PLDs are indicated by the black arrows on the left. When the PLDs are the same, the excursion sizes, indicated by the black arrows on the right, differ by the amount represented by S, indicated by the open arrow, also on the right.
It can be seen that high pitch levels on the toplines of the stimuli turn out to be roughly equal when the PLDs are equal (open ellipsis on the left and hatched ellipsis on the right). Since the task of matching PLDs may have been quite complex, an alternative would be that listeners adjusted the indicated high pitch levels, viz. in the *first* syllable, i.e. before the fall, and in the *third* syllable, i.e. after the rise, so as to make them equal[13]. Note that the first and the last syllable are unaccented.

A linguistic reason why matching high pitch levels may have been a valid strategy can be found in the experiments by Kutik *et al.* (1983). They found that listeners matched the toplines of a main clause before and after a parenthetical clause which had different lengths. In our experiments, listeners may have interpreted the stimuli in a pair as being the parts of one single utterance containing two pitch-accented syllables, especially in the case of a rise followed by a fall. The pause would than be the parenthetical.
Adjusting high pitch levels so as to make them equal may have been even more plausible as to be a strategy due to the actual experimental setup. In the present experiments, differences in prominence were brought about by changing the height of the topline relative to the fixed baseline, because it is known from perception experiments that variation of high pitch levels is perceptually more salient for the perception of intonation than variation of low pitch levels (e.g. Sluijter and Terken, 1993). Since the excursion size was controlled by changing the height of the topline relative to the fixed baseline, changes in the height of high pitch levels were the only cue to changes in prominence. This might suggest that subjects had adjusted pitch levels on the toplines to become equal.

[13] The hypothesis that listeners focused on the high pitch level in the first syllable of the stimulus containing the fall and on the high pitch level in the accented second syllable of the stimulus containing the rise does not lead to the right predictions. The predicted shift would then be equal to $S = dT$. That means that it would be only about half as large as the shift found in experiment I and in the experiments by Hermes and Rump (1994).

In the current experiment, these two strategies, i.e. the matching of PLDs and the matching of high pitch levels, were not conflicting. Actually, both strategies lead to the same result. The next experiment was designed in such a way that following one strategy would lead to another result than following the other strategy.

The strategy of equating high pitch levels can be formulated quantitatively in the following way. As shown in Fig. 4, it is most likely that the listeners chose the high pitch level in the first syllable of the stimulus containing the fall and the high pitch level in the third syllable of the stimulus containing the rise. The shift S is then determined by the rate of declination, d, and by the time interval between the relevant pitch levels in the first and the third syllable, T', or

$$S = dT' \qquad (3)$$

If we want to predict the shifts that would have resulted in experiment 1, we may assume, as we did in the PLD model, that the time interval between relevant pitch levels is of the same order as the time interval between the onsets of the vowels containing the relevant pitch levels. This time interval between the first and the third syllable will be referred to as V'. Measurements showed that V' was about 470 ms, which turned out to be nearly twice the time interval between the vowel onsets of the first and the second syllable (V: 220 ms). The newly predicted shifts based on the matching of high pitch levels will then be

$$S = dV', \qquad (4)$$

instead of $S = 2dV$, but since $2V$ and V' were almost equal, the two predictions are only marginally different.

The predictions derived from Eq. 4 are shown in Table II, together with the shifts measured in experiment 1 and the shifts predicted by the PLD model (Eq. 2).

It can be seen that the observed shifts are between the predictions derived from the PLD model and the predictions derived from Eq. 4, at least in the case of the short and original stimuli. In the case of the long stimuli, however, both predicted values of the shift are too large. It is not clear why this has been the case. For both strategies, *ad hoc* explanations may be given:

Table II. *Shifts (S) predicted by the PLD (Eq. 2) and H (Eq. 4) models, and the mean observed shifts with their standard deviations in experiment 1. Measures in E.*

		stimulus duration	
	short	original	long
pred. (Eq. 2)	0.300	0.440	0.700
pred. (Eq. 4)	0.340	0.470	0.690
obs.	0.334	0.453	0.540
sd	0.056	0.093	0.055

- In terms of the PLD model, as mentioned before, it may be assumed that the contribution of pitch to prominence is not proportional to the pitch at points fixed relative to the vowel *onsets* but that a pitch level early in the vowel of the accented nucleus is compared with a pitch level late in the vowel of the preceding syllable. In other words, *T* increases with an increase of *V*, but the increment is somewhat smaller.
- In terms of the pitch-matching strategy, this means that the listeners matched a lower than predicted pitch of the third syllable to the pitch of the first. It may be assumed that, due to its relatively long duration, the high pitch level of the accented *second* syllable had become a better candidate for the match with the high pitch level in the first syllable than the high pitch level in the unaccented third syllable. In other words, *T* changes due to the fact that a different pitch level becomes more relevant.

It is thus argued that listeners may have adopted the strategy of matching high pitch levels rather than matching pitch-level differences. We will refer to this new strategy as the High-level (H) model[14]. In experiment 2, the PLD model and the H model were compared by using stimuli with different baselines.

[14] Note that, in this extreme form, the H model claims that listeners do not use any information about low pitch levels when adjusting prominences to be equal. From earlier experiments with stimuli in different registers (Hermes and Van Gestel, 1991; Hermes and Rump, 1994), however, it is obvious that listeners do use information about baselines when determining the prominence lent by pitch movements.

4.3 EXPERIMENT 2

From the discussion of experiment 1 it has become clear that the PLD model and the H model both make rather similar predictions if the low pitch levels are on baselines with equal endpoints. According to the PLD model, pitch levels on the baseline and pitch levels on the topline are equally important, because the listeners need information about low and high pitch levels in order to be able to determine the PLDs. According to the H model, however, only pitch levels on the toplines in the test and comparison stimuli determine the finally adjusted excursion sizes. If the baselines in the test and comparison stimuli are different, the PLD model and the H model make different predictions about the resulting shifts. In experiment 2 we tested whether low pitch levels contributed to the perceived degree of prominence.

4.3.1 Method

4.3.1.1 Stimuli, procedure, and subjects. As in experiment 1, the stimuli consisted of PSOLA-manipulated versions of the nonsense utterance /ma'mama/. The prominence lent by a rise was compared with the prominence lent by a fall. The stimuli were resynthesized in the same low register. The slopes of the declination lines were kept constant at 0.7 E/s. Unlike in experiment 1, the end frequencies of the baseline in the test and comparison stimuli were chosen to be different. Two new excursion-size continua were constructed, in which the end frequency of the baseline was shifted up or down by 11 Hz, about 0.33 E or 2.4 semitones, relative to the formerly used end frequency of 75 Hz, resulting in end frequencies of 64 Hz and 86 Hz. (Since the baselines ran parallel in the ERB-rate frequency domain, the difference of 11 Hz between the end frequencies amounted to a difference of 0.33 E between the baselines, which means that differences between the begin frequencies were slightly larger than 11 Hz.) In the remainder of the paper we will refer to stimuli with rises as R and stimuli with falls as F. The end frequency of the baseline will be indicated by specifying the Hz value in subscript after the pitch movement. For example, R_{75} means that the (virtual) end frequency of the baseline in the stimulus containing the rise was 75 Hz.
Stimuli with a shifted baseline were always compared with stimuli having the

original baseline with the end frequency of 75 Hz. If the listeners had attended mainly to the high pitch levels in the unaccented syllables before the fall and after the rise, the resulting shift would have been different from the shift predicted by the PLD model.

If the relevant high pitch levels, in the first syllable, before the fall, and in the third syllable, after the rise, had been made equal, the predicted shift would have been

$$S = dV' \pm 0.33. \qquad (5)$$

Since dV' is equal to 0.33 E under the present stimulus conditions, the H model (Eq. 5) predicts shifts of either 0.00 or 0.66 E. The predictions by the H model are illustrated in Fig. 5.

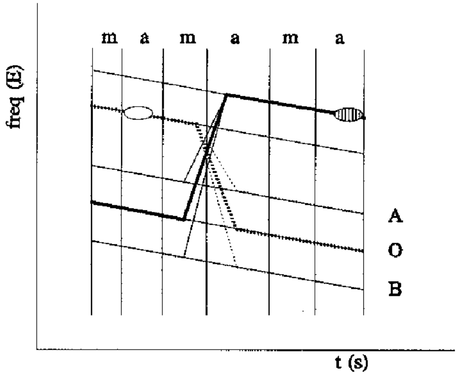


Figure 5. *Illustration of the predicted excursion sizes on the assumption of the H model. Baseline O is the original baseline. Baseline A results when baseline O is shifted 0.33 E upwards, baseline B results when baseline O is shifted 0.33 E downwards. The H model predicts that the excursion sizes of the rise and the fall are equal if the fall has baseline O or B and the rise has baseline A or O, respectively. The H model predicts that the excursion sizes of the rise and the fall differ 0.66 E from each other if the fall has baseline O or A and the rise has baseline B or O, respectively. The PLD model predicts that the excursion size of the rise will be 0.33 E larger than that of the fall, independently of the different baseline combinations for the rise and the fall.*

Equation 5 predicts a shift zero when a rise is compared with a fall having a lower baseline, since high pitch levels are equal when excursion sizes are equal (comparisons between a rise with baseline A and a fall with baseline O, and between a rise with baseline O and a fall with baseline B). Equation 5 predicts a shift of 0.66 E when a rise is compared with a fall having a higher baseline (comparisons between a rise with baseline O and a fall with baseline A and between a rise with baseline B and a fall with baseline O).

The newly predicted shifts are either zero or more than twice the size of the shift predicted by the PLD model, depending on the heights of the baselines in the test and comparison stimuli. The shift predicted by the PLD model was assumed to be independent of the different baselines of the stimuli, and to depend only on the height of the topline relative to the baseline within each individual stimulus. The shift would therefore always be *2dT* or about 0.31 E.
The experimental procedure was identical to the procedure used in experiment 1. The total number of adjustments per subject was 64 (4 end-frequency combinations: 64-75, 75-64, 86-75, and 75-86; 2 pitch movements on the test stimulus; 4 excursion sizes of the test stimulus; 2 starting points in the excursion-size continuum). The order of presentation was randomized.
Nine subjects participated in the experiment. All except one had taken part in the previous experiments. The same consistency criterion was applied as in experiment 1. A subject's data were included in the analysis if the score was consistent for at least three of the four end-frequency combinations.

4.3.2 Results and discussion

The data of two subjects had to be omitted from the analysis because they did not show the required consistency. The same subject as in experiment 1 had made musical intervals equal. His data were therefore also omitted. The analysis of the results of the six remaining subjects showed that the effect of the starting points in the excursion size continuum (hysteresis) was much greater in the present experiment than in experiment 1: 0.25 E (experiment 1: 0.14 E). It was of the same order of magnitude for all comparisons.
The results are shown in Fig. 6. As in Fig. 3, the adjusted excursion sizes were averaged across the starting points and across the subjects. Panels (a)

through (d) show the results obtained for the four comparisons. The open circles having horizontal standard-deviation bars again refer to the adjusted excursions of the rise, the black circles having vertical standard-deviation bars refer to the adjusted excursions of the fall. The solid line in each panel indicates the position of the measurement points if equal excursion sizes were to lend equal prominence. The dashed line in each panel indicates the shift predicted by the PLD model. The distance between the solid and the dashed line measured perpendicularly to the horizontal axis is therefore 0.31 E, and is the same in all panels.

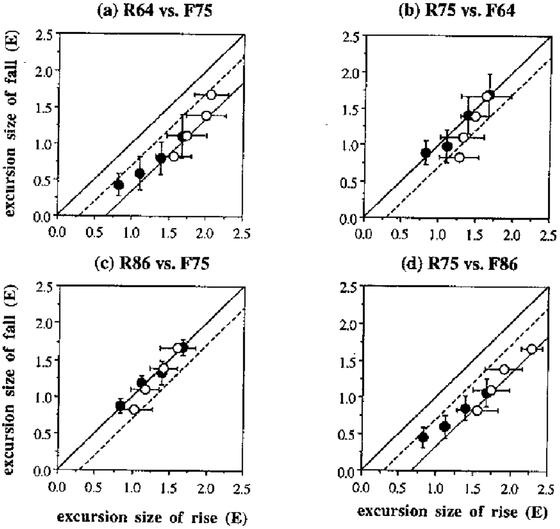


Figure 6. *The results of the six consistent subjects who participated in experiment 2. In (a) the results are shown for the comparison of R_{64} vs. F_{75}, in (b) for R_{75} vs. F_{64}, in (c) for R_{86} vs. F_{75}, and in (d) for R_{75} vs. F_{86}. The diagonal indicates equal excursion sizes, the dashed line indicates the values predicted by the PLD model, the dotted line indicates the values predicted by the H model [the dotted line and the diagonal coincide in panels (b) and (c)]. As in Fig. 3 the test stimulus is on the horizontal axis when falls are adjusted and on the vertical axis when rises are adjusted.*

The dotted lines indicate the shifts predicted by the H model. Therefore, the distance between the solid and the dotted lines depends on the combination of the baselines in the pair of stimuli, the shifts being about zero in panels (b) and (c), in which the diagonal and the dotted line coincide, and 0.66 E in panels (a) and (d).

As in experiment 1, the regression lines for the adjusted excursion sizes of rises and falls were slightly different due to the effect of a trend towards the average explained above.

The observed shifts for the individual subjects are presented in Table III. The results of all the subjects showed the same trends. A repeated-measures ANOVA across subjects showed that different end frequencies of the baselines had a significant effect on *S* ($F_{(3,15)} = 112.9$, $p < 0.001$). Mean shifts and their standard deviations were calculated from the distances between the individual data points and the diagonal, measured perpendicularly to the horizontal axis, in each of the panels in Fig. 6.

Table III. *Individual shifts (S) in experiment 2, together with the shifts predicted by the PLD (Eq. 2) and H (Eq. 5) models, and the mean observed shifts with their standard deviations. Measures in number of ERBs.*

	Comparison			
	$R_{64}F_{75}$	$R_{75}F_{64}$	$R_{86}F_{75}$	$R_{75}F_{86}$
Subj.				
1	0.286	0.043	-0.043	0.373
2	0.833	0.234	0.200	0.799
3	0.582	0.226	0.095	0.599
4	0.547	0.052	0.017	0.582
5	0.573	0.069	-0.052	0.512
6	0.503	-0.078	-0.139	0.495
pred. (Eq. 2)	0.330	0.330	0.330	0.330
pred. (Eq. 5)	0.660	0.000	0.000	0.660
obs.	0.554	0.091	0.013	0.560
sd	0.112	0.169	0.087	0.100

The shifts predicted by the PLD model and the H model are also displayed. As predicted by the H model, the shifts were relatively large in the comparisons in which the rise had a lower baseline than the fall (R_{64} vs. F_{75} and R_{75} vs. F_{86}), and the shifts were relatively small in the comparisons in which the rise had a higher baseline than the fall (R_{75} vs. F_{64} and R_{86} vs. F_{75}).

The results were found to be closer to the values predicted by the H model (Eq. 5) than to the values predicted by the PLD model (Eq. 2). This suggests that matching high pitch levels played a significant role when subjects adjusted prominence. But the fact that the measurements were between the two predictions indicates that the influence of the low pitch levels was significant, too.

4.4 GENERAL DISCUSSION AND CONCLUSIONS

In experiment 1 it was shown that a change in the time interval between successive vowel onsets had a significant influence on the adjusted excursion sizes of the rise and the fall. The shift, indicating the difference between the excursion sizes of a rise and a fall lending equal prominence, decreased when the stimuli were shortened, and increased when the stimuli were lengthened. When these results are interpreted in terms of the PLD model this means that a change in the time interval between vowel onsets indeed resulted in a change in the time interval between pitch levels which were assumed to be relevant for the determination of the PLDs. The hypothesis that T is a fixed time interval is therefore not supported.

Although the measured shifts were close to the values predicted by the PLD model, they did not fully support the model. An alternative explanation was proposed, to the effect that listeners attended to high pitch levels located in the unaccented syllables of the stimuli when adjusting prominence so as to make it equal (H model). The reanalysis of the results of experiment 1 showed that the predictions made by the H model matched the results fairly well, at least as far as the short and original stimulus durations are considered.

In experiment 2, we found that the listeners tended to attend to high pitch levels, making them about equal when adjusting prominence. The role of low pitch levels turned out to be significant, too, which means that both strategies, the one following the PLD model and the one following the H model, played a role when the subjects performed the adjustment task.

In summary, if the baselines are equal, the strategy following the H model leads to the same adjustments as the strategy following the PLD model. If the baselines are different, however, the strategies lead to different adjustments which makes the task more difficult. This is illustrated by larger effect of hysteresis in the case of slightly different baselines, as in experiment 2.

It was already mentioned in Footnote 5 that it is obvious that the H model does not apply when stimuli are in different registers, as in experiment 1 by Hermes and Rump. In those cases the low pitch levels in each individual stimulus necessarily provide some kind of reference for judging the height of the high pitch levels. As mentioned earlier, however, the standard deviations under the different-register conditions were much larger than under the equal-register conditions. But also here, the fall significantly lent more prominence than the rise and the rise-fall.

In their second experiment, Hermes and Rump used stimuli which were in the same register. They found that falls lend more prominence than rises and rise-falls. Furthermore, varying the timing of the falls did not affect the prominence they lent as compared to the prominence lent by a rise-fall. Early rises lent the same prominence as a rise-fall and for rises starting later than the vowel onsets, the prominence was reduced or became difficult to compare to the prominence lent by a rise-fall. In these experiments, comparisons always included a rise-fall with only one high pitch level on the topline, viz. the level of the vowel of the accented syllable. It is difficult to say how the strategy as given by the H model will have affected these results.

In conclusion, the role of high and low pitch levels in the perception of prominence of pitch-accented syllables is not clear yet. Under the same-register condition, high pitch levels seem to contribute more to prominence than low pitch levels do, although the role of the latter is still significant. As mentioned before, in the case of stimuli in the same register, the strategy of matching high pitch levels is a plausible strategy, too, according to Kutik *et al.*

This strategy, according to Kutik, then implicates that the two stimuli presented as a pair are considered by the subjects as constituents of one large phrase. Redoing the experiments with stimuli in different registers may therefore be more conclusive, although it has to be taken into account that standard deviations will be much larger in that case. A reason for the larger standard deviations seems to be the fact that the strategy of matching high pitch levels, i.e. making use of an additional cue for equal prominence, is no longer possible.

5 PROMINENCE OF TWO PITCH-ACCENTED SYLLABLES IN ONE UTTERANCE: EFFECTS OF DECLINATION[15]

ABSTRACT

Two experiments were conducted to investigate the effect of declination on the relative perceptual prominence of two accented syllables within the same 7-syllable utterance. The maxima of the two pitch peaks (p1, p2) were varied for each of three rates of exponential pitch declination, keeping the terminal pitch constant. Extending earlier work by Terken (1991) with low-pitched reiterant speech, a meaningful Dutch sentence was employed here in both female and male pitch ranges. Listeners judged which pitch-accented syllable, s1 or s2, was more prominent (experiment 1a) or adjusted the pitch peak maximum p2 on s2 to match s1 in prominence (experiments 1b and 2). The present study extends previous findings on the perception of the relative prominence of two pitch accents within a single utterance, and it shows that the situation is somewhat more complex than predicted by existing prominence models. By varying declination rate and the value of p1 in a semi-independent fashion, the effects of these two factors on prominence perception could be assessed. The judgement and adjustment methods gave the same results, as did the use of female and male pitch ranges. The effect of declination rate was surprisingly small in terms of peak maxima, but it was large in terms of excursion sizes from the actual baseline. The implications of these results for models of prominence perception are discussed.

[15] The data of experiment 1 were presented at the 126th meeting of the Acoustical Society of America, Denver, Colorado, October 1993, as B.H. Repp, H.H. Rump, and J.M.B. Terken: "Relative perceptual prominence of fundamental frequency peaks in the presence of declination," J. Acoust. Soc. Am. **94** (1993), 1880. They were published under the same title in the Annual Progress Report, Institute for Perception Research/IPO, **28** (1993), pp. 59-62. The data will also appear in a paper by C. Gussenhoven, B.H. Repp, A. Rietveld, H.H. Rump, and J.M.B. Terken: "The perceptual prominence of fundamental frequency peaks" (submitted for publication in JASA).

5.1 INTRODUCTION

It has often been observed that the periodicity of the speech signal, which is perceived as pitch, tends to decrease in the course of a (declarative) utterance. This phenomenon of declination of the pitch contour occurs in many (perhaps all) languages, and it has been thoroughly studied in Dutch (see 't Hart, Collier, and Cohen, 1990). The rules for intonation synthesis developed by 't Hart *et al.* include a "declination line" or baseline, i.e. a lower bound to periodicity which decreases exponentially over time within an utterance and is explicitly apparent in the pitch of unaccented syllables. A suitable slope of this baseline for synthetic Dutch speech is -11/t+1.5 semitones per second (st/s), where t is the duration of the utterance (up to 5 s).

In this study we investigate the influence of declination on the perception of the relative prominence of two accented syllables within the same utterance. It seems a plausible hypothesis that the prominence of pitch peaks superimposed on a declining baseline would be perceived relative to that baseline (i.e., relative to the low pitch levels in adjacent unaccented syllables) rather than in terms of absolute peak height, particularly when the task is to evaluate prominence rather than pitch as such. Therefore, when the rate of declination across the unaccented syllables is varied, while holding absolute peak heights constant, the relative prominence of accented syllables should change.

Several previous studies with stimuli containing more than one pitch-accented syllable in the same utterance have explored related questions, though they have not separated the effects of declination rate from those of peak height. All these studies used reiterant speech stimuli (of the form /mamama.../) to avoid effects of phonetic and linguistic structure, and all required listeners to judge the relation between two pitch peaks superimposed on some baseline. The utterances were always 6-8 syllables in length, spoken in a male voice. The height of the peaks was manipulated using LPC analysis and resynthesis techniques.

In the earliest of these studies, Pierrehumbert (1979) employed modified natural speech (American English) with an unspecified (but small) amount of baseline declination. Only the maxima of the pitch peaks on the first and second accented syllables (in our terms p1 and p2, and s1 and s2, respectively) were varied. The subjects were asked to judge whether s1 or s2 seemed to have a higher pitch. The subjects' responses indicated that, when the value of p1 was large (corresponding to an excursion size of 11 st above the baseline), it was perceived as equal in pitch to a value of p2 that in fact was about 1 st smaller. This was attributed to a perceptual compensation for expected "topline" declination, where the topline describes the difference between the pitch-peak maxima p1 and p2. To the extent that such a difference is commonly found in natural speech (cf. 't Hart *et al.*, 1990), it is reasonable to assume that listeners expect it to occur in experimental utterances as well.

In a subsequent study with Dutch listeners, Leroy (1984) systematically varied baseline declination while also varying the value of p2, keeping the excursion size of the pitch peak on s1 constant (5 st). In one condition, there was no baseline declination at all, but subjects nevertheless showed a compensation effect: p1 was judged as equal in pitch to p2 that was about 1 st lower. When the standard Dutch synthesis rate of baseline declination was present in the stimuli, the difference between p1 and p2 judged to be equal in pitch increased to 1.8 st, and an even steeper declination line resulted in a further small increase in the p1-p2 difference to 2.1 st. It appears from these results that when the slope of the baseline is raised, the expected slope of the topline does not increase at the same rate: once there is substantial declination, the expected topline declines less than the baseline. This is not unreasonable, as a high pitch at utterance onset constrains the excursion size on s1, whereas no such constraint is present on s2 later in the utterance.

These early studies may have underestimated the p1-p2 difference because they asked for pitch judgments rather than prominence judgments. Pitch judgments may be only an indirect reflection of perceived prominence. Gussenhoven and Rietveld (1988) were the first to ask their subjects to rate the *degree of prominence* (rather than the pitch height as such) of s1 and of s2 on a 10-point scale; however, they obtained these judgments for each peak separately. Even so, s2 received higher ratings than s1 when their absolute peak maxima were about the same. Gussenhoven and Rietveld also collected production data, and it appeared that the judged p1-p2 difference exceeded the difference due to baseline declination in production; the authors attributed this to an expectation of final lowering in addition to topline declination. The declination rate in their synthetic utterances, however, was much smaller than is typical for Dutch ('t Hart *et al.*, 1990), so the results are more similar to those obtained with a flat baseline. (See also the discussion between Terken, 1989, and Gussenhoven and Rietveld, 1989.) In addition, Terken (1994) has shown that the lowering effect occurs to the same extent when the adjusted pitch peak is not the last one in the utterance but is followed by another one.

The most thorough investigation to date of the relative perceptual prominence of two accent peaks in one utterance was conducted by Terken (1991). In one experiment he used stimuli without any baseline declination, varying the value of p1. In a second experiment, he covaried the rate of baseline declination and the value of p1, such that the excursion size of the peak on s1 was always about 6.2 st above the baseline. (Since the baseline always ended at the same low pitch value, the pitch level of unaccented syllables covaried with declination rate throughout the utterance.) Terken also compared two types of instruction: to match s1 and s2 in either pitch or prominence. Using an interactive computer set-up, the subjects adjusted the value of p2 until s2 matched s1 according to the stated criterion.

When there was no declination in the stimuli, the difference between p1 and p2 judged to be equivalent was negligibly small with pitch-matching instructions, probably because the listeners were experienced speech researchers who listened analytically. With prominence-matching instructions, a small difference (p2 < p1) emerged at excursion sizes greater than 8.5 st. When the declination rate varied from 0 to 5 st/s, the adjusted values of p2 again tended to be smaller than those of p1 with pitch-matching instructions, and this difference tended to increase with p1. With prominence-matching instructions, however, the p1-p2 difference was much larger and increased systematically with higher values of p1. The effect was present in all subjects, though its magnitude varied. The average results did not support the hypothesis that equal prominence corresponds to an equal distance from the baseline in semitones: rather than being constant at about 6.2 st (the average distance of p1 from the baseline, d1), adjusted values of d2 increased with higher values of p1 from 5.7 to 8.3 st. The Hertz scale or the ERB-rate scale, proposed by Hermes and Van Gestel (1991) as the most appropriate scale for perceived prominence, did not yield equal distances either.

Ladd (1993), in commenting on Terken's results, pointed out that they could be accounted for by assuming that the perceived prominence is determined not by peak height relative to the actual baseline but with reference to a mental reference line of fixed slope whose absolute height varies somewhat with the rate of declination in the stimulus. However, Terken (1993) countered that it is not clear how the listener would determine the slope and current value of this reference line, and that Ladd's model generates some implausible predictions for conditions outside the range of Terken's (1991) data. He instead proposed a different model which incorporates the bottom (lowest value) of a speaker's pitch range as a fixed reference value. This is an individually stable point usually reached at the end of an utterance (cf. Maeda, 1976; Liberman and Pierrehumbert, 1984; 't Hart *et al.*, 1990). A problem with this reference is that listeners have it available only if they have heard the speaker before.

The aim of this study was to provide additional data that may help refining models of prominence perception. Terken's design was limited in that it varied the value of p1 and the declination rate concurrently. This made it impossible to gauge the relative contributions of these two factors to perceived s1 prominence. In this study, the value of p1 and the declination line were varied semi-independently, so that for each declination rate several values of p1 were tested. Terken and his predecessors also focused exclusively on male speech; that is, the end point of the declination line was always at a relatively low frequency. This study attempted to correct this situation by starting out with female speech (experiment 1), though later a male speech condition was added for comparison (experiment 2). Instead of the customary reiterant speech, a simple meaningful sentence was used. This was expected to help subjects maintain a linguistic frame of reference, at the risk of introducing other confounding variables. In experiment 1, two tasks were compared: a single-trial relative prominence judgment task (henceforth, ''judgment task'') and an adjustment task, as used by Terken (1991). The judgment task was particularly thought to discourage a literal pitch-matching strategy. Finally, a group of relatively naive subjects (i.e., not including any speech researchers) was used. In Terken's (1991) study there seemed to be a tendency for the most experienced listeners to show the smallest effects. The segmental homogeneity of his stimuli, his adjustment paradigm, and his listeners' experience all encouraged an analytic listening strategy that may have counteracted the adoption of the ''linguistic mode'' most suited to prominence judgments.

In summary, this study addressed the following main questions:

(1) Given the various changes in methodology, will the basic finding be

replicated, i.e. that the value of pitch peak maximum p2 is smaller than that of p1 when the associated pitch-accented syllables s2 and s1 are judged to be equally prominent?

(2) Does this discrepancy increase with the value of p1, as it did in Terken's (1991) study?

(3) Is there an effect of declination rate above and beyond the effect of p1 value?

(4) What model of prominence perception accounts best for the results obtained?

5.2 EXPERIMENT 1

5.2.1 Method

5.2.1.1 Materials. A female Dutch speaker spoke the sentence ''A'manda gaat naar 'Malta'' (''A'manda is going to 'Malta'') in a neutral fashion, i.e. without special emphasis on either ''Amanda'' or ''Malta'' (Terken had asked his subjects to keep this sentence in mind when judging the reiterant utterances). The speech was recorded digitally in a recording studio using high-quality equipment. Subsequently it was put into a computer, digitized at 20 kHz, and low-pass filtered at 10 kHz. Phonetic segment boundaries were determined using a waveform editing program. The total utterance was about 1,635 ms in duration; the critical /mɑn/ and /mɑl/ syllables had durations of 291 ms (/m/: 87 ms; /ɑn/: 204 ms) and 268 ms (/m/: 106 ms; /ɑl/: 162 ms), respectively. The utterance was then LPC-coded (24 coefficients, 10 ms frames), and the pitch contour was determined using the method of subharmonic summation (Hermes, 1988).

The pitch values of the unaccented syllables were matched to a single straight line which happened to decline at a rate of 2.4 st/s. In order to make sure that pitch manipulations would have comparable effects on the first and second pitch peaks, the accent-lending pitch movements were given equal temporal alignments with the segmental structure of the utterance, keeping the original form of the first pitch peak as it was. As a result of this, rises started 30 ms before the vowel in the accented syllable and lasted 110 ms. Falls started after a short declining plateau of 70 ms and lasted 130 ms (these values were very close to the ones in Terken, 1991, following the standard rules of Dutch synthesis from 't Hart *et al.*, 1990). By coincidence, the time interval between the pitch peak maxima was identical to that in Terken, 1991, i.e. 900 ms. The original and the stylized pitch contours are shown in Figure 1. There were two gaps in the pitch contour due to voiceless segments: the voiceless velar fricative /X/ at the beginning of "gaat" and the /t/ in "Malta". (The final /t/ in "gaat" was assimilated to the following nasal and exhibited continuous voicing.) Even though the /t/ closure occurred during the fall of the second peak, there was little reason to believe that this would affect the perceived prominence of s2.

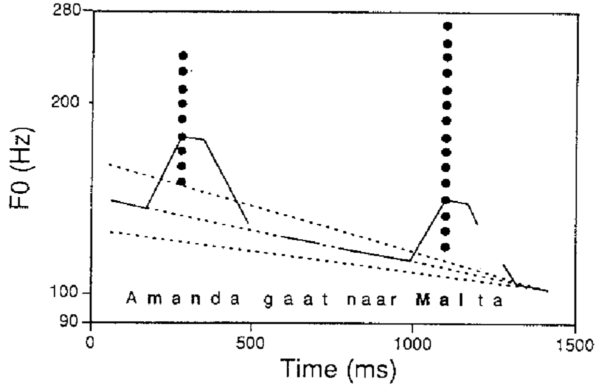


Figure 1. *Schematic illustration of stimulus parameters in experiment 1.*

With the stylized version as the starting point, a set of 105 pitch contours was constructed by combining 3 declination rates, 5 values of p1, and 7 values of p2. The combinations were not strictly orthogonal because the range of p1 values necessarily had to covary to some extent with the height of the baseline, and the range of p2 values had to be approximately centered around each p1 value to yield a good estimate of the point of equal prominence. Therefore, 9 p1 values were used which were assigned to the 3 declination lines in 3 overlapping sets of 5, and 15 p2 values were assigned to the 9 p1 values in 9 overlapping sets of 7. The values of p1 ranged from 247 to 392 Hz in steps of approximately 1 st, and p2 values ranged from 198 to 444 Hz, also with a 1 st step size. The three declination rates were 2.4 (original), 3.7, and

4.9 st/s. The lowest pitch value, reached at the end of each declination line, was 164 Hz. These parameters are illustrated in Figure 1. The precise stimulus design can be derived from Table I below.

For the judgment task (experiment 1a), the 105 stimuli were randomized with interstimulus intervals of about 4 s. A longer interval occurred after the 35th and the 70th trial. Five different randomizations were created, separated by longer pauses. Each of these blocks lasted about 10 minutes. The five test blocks were preceded by a practice sequence containing, in random order, the 30 stimuli in which, for each p1 value, p2 had one or the other extreme value in its 7-value range (i.e. the first and last stimuli in the rows of Table I below). The test sequences were recorded on digital tape.

The adjustment task (experiment 1b) used an interactive computer program and thus did not require recording of stimulus sequences.

5.2.1.2 Subjects and procedure. The subjects were 7 members of the research staff at IPO, 4 women and 3 men, all native speakers of Dutch, who volunteered to participate. None of them is involved in speech research.

The judgment task, including instructions and a short break after the third block, took a little over one hour. Subjects were tested in two groups in a quiet listening room. The stimulus sequences were presented over earphones at a comfortable intensity level. Written instructions were presented in English, which is spoken fluently by all participants. The subjects were asked to judge for each stimulus which of the two words, ''Amanda'' or ''Malta'', was given more emphasis by the speaker. (The instructions pointed out that this was the same as judging the relative prominence of the accented /mɑn/ and /mɑl/ syllables.) The responses were made by writing either ''1'' or ''2'' on a prepared answer sheet. A forced choice was required on every trial.

About two months later, the same subjects took part in the adjustment task, which lasted about 30 minutes. Subjects were tested individually in a quiet room. Instructions similar to those in the judgment task were given in Dutch. Subjects sat in front of a computer terminal and listened to the stimuli over earphones. They completed 60 adjustment runs, 4 for each of the 15 combinations of p1 and declination rate, arranged in 4 blocks of 15 runs each. In each run there were 7 available values of p2 (cf. Table I below); half the runs started with the lowest value of p2, the other half with the highest. Using the computer keyboard, the subject could increase or decrease the value of p2 and listen repeatedly to the utterance until a satisfactory match of s2 with s1 was achieved. That match was registered by the computer. The adjustments were subsequently averaged over the 4 replications.

5.2.2 Results and discussion

For the 105 test stimuli in the judgment task the number of "1" responses was tallied, added up across subjects, and converted into percentages. These percentages are shown in Table I. Although some subjects were not totally confident even at the extremes of the p2 ranges, an orderly progression from "1" to "2" judgments can be seen in each of the 15 rows of the table.

Table I. *Percentages of "1" responses to the 105 test stimuli (experiment 1).*

Decl. (st/s)	p1 (Hz)	p2 (Hz)														
		198	211	222	235	250	263	278	294	313	333	351	370	392	417	444
2.4	247	94	77	80	66	54	26	11								
	263		97	91	71	69	31	20	9							
	278			86	86	71	63	29	9	3						
	294				97	77	46	43	34	0	9					
	313					91	74	54	43	17	3	6				
3.7	278			94	97	80	63	49	17	11						
	294				89	80	77	51	34	6	9					
	313					94	91	77	46	26	3	0				
	333						91	83	66	31	11	3	14			
	351							97	77	49	34	6	3	3		
4.9	313					91	83	89	60	14	26	6				
	333						94	91	63	66	20	9	3			
	351							91	80	57	37	20	9	6		
	370								94	80	63	26	6	6	0	
	392									89	80	54	17	3	0	0

Probit analysis (Finney, 1971) was used to estimate the 50% crossover point from "1" to "2" responses within each of the 15 series of 7 p2 values, individually for each subject. These estimates represent the points of subjective equal prominence of s1 and s2. Figure 2a shows a plot of these estimates, averaged across the 7 subjects. Figure 2b shows the results of the adjustment task, which yielded such estimates directly. Analysis of the data showed that there was an experimental bias which occurred only in the adjustment task: when p2 started with its highest value, adjustments were higher than when it started with its lowest value. In order to cancel this effect (which we refer to as "hysteresis" and which was also observed in the experiments described in chapter 3), the data were averaged over both types of adjustment runs.

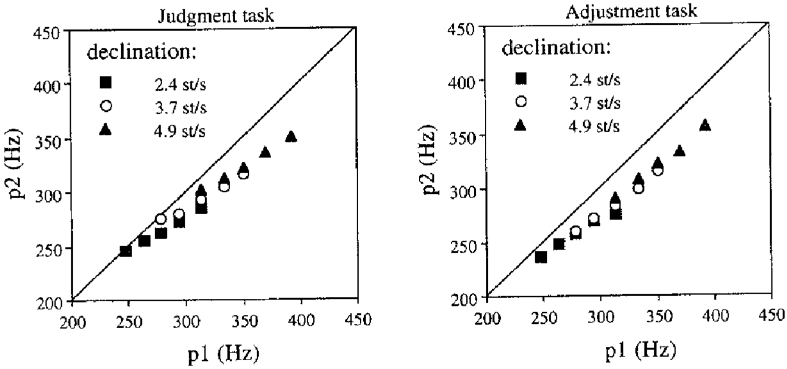


Figure 2. *Equal-prominence p2 values as a function of p1 value in the two tasks of experiment 1.*

Despite some variability at the individual level, the overall results were remarkably orderly. The average data show the expected p1-p2 difference: the points are all below the diagonal, which means that the value of p2 was smaller than that of p1 when s1 and s2 were judged to be equal in prominence. All 7 subjects showed this overall effect in both tasks.

Figure 2 also indicates that the discrepancy between p1 and p2 increased with p1. The consistency of this effect was assessed by fitting straight lines to the 5 data points for each declination condition, separately for each subject (there were no systematic nonlinearities in any subject's data), and by examining the slopes of these regression lines. They are shown in Table II, and it is clear that nearly all are less than 1. Thus the increase in the p1-p2 difference with p1 was also statistically reliable. The slopes tended to be steeper in the adjustment task, but this task difference fell short of significance [$F_{(1,6)} = 5.6$, $p < .06$].

Table II. *Slopes of straight lines describing the p1-p2 equal-prominence relationship in each declination condition in the two tasks (experiment 1).*

Subject	Judgment task			Adjustment task		
	2.4 st/s	3.7 st/s	4.9 st/s	2.4 st/s	3.7 st/s	4.9 st/s
#1	0.31	0.33	0.28	0.49	0.68	0.69
#2	0.89	0.71	0.76	0.62	0.85	0.83
#3	0.58	0.50	0.36	0.28	0.41	0.67
#4	0.62	0.77	0.77	0.63	1.09	1.05
#5	0.57	0.70	0.72	0.61	0.59	0.79
#6	0.40	0.52	0.66	0.66	0.70	0.73
#7	0.75	0.62	0.77	1.00	0.97	0.83
Average	0.59	0.59	0.62	0.61	0.76	0.80

The third and principal effect of interest in the data, that of declination, was less pronounced. In Figure 2, a slight staggering of the data points for the three declination conditions may be observed. The significance of this effect was tested in three ANOVAs on the p2 estimates: for the two sets of p1 values that were shared by two declination conditions, and for the single p1 value that was shared by all three. The declination effect was significant in all three analyses [low vs. medium declination rate: $F_{(1,6)} = 13.4$, $p < .02$; medium vs. high declination rate: $F_{(1,6)} = 56.9$, $p < .0004$; all three: $F_{(2,12)} = 17.6$, $p < .0004$], and there was no significant interaction with task or with p1 value. Of course, there were highly significant main effects of p1 in the first two analyses, but the interactions of that effect with task were nonsignificant. The ANOVA on the slopes of the lines fitted to the data points (Table II) also revealed a significant effect of declination [$F_{(2,12)} = 4.3$, $p < .04$], indicating that they got steeper as the declination rate increased. In other words, the increase in the p1-p2 difference with p1 was more pronounced with low than with high declination rates. This effect did not interact with task either. The judgment and adjustment tasks thus gave statistically equivalent results. Table II suggests, nevertheless, that the slope differences were essentially restricted to the adjustment task.

It may be asked now whether this pattern of results is consistent with the recent models of either Ladd (1993) or Terken (1993). Ladd's model is a simple model which he proposed only to provide an alternative explanation of Terken's (1991) data. According to this model, listeners judge peaks with respect to an internal reference line of fixed slope. This declination line is ''anchored'' on the interpolated baseline below p2 and thus moves up or down slightly as the actual baseline declination changes. Syllables are assumed to be perceived as equally prominent when their pitch movements have equal excursion sizes, i.e. when their pitch maxima are equidistant from the reference line. Parallel translation of the reference line up or down the frequency axis leaves the maxima equidistant, no matter what metric (Hz, E, or st) is employed. Thus, declination rate should have no effect on relative prominence. This model is inconsistent with the declination effect shown in Figure 2. However, the size of that effect, although it is statistically significant, is rather small, so the model may still be correct to a first approximation.

The model fails more seriously in predicting another aspect of the data: an increase in p1 should cause an equal increase in matched p2 in terms of their distances from the reference line. If the wrong reference line (i.e. the actual baseline, in Ladd's view) is employed for calculating distances, then d1 and d2 will be unequal for peaks lending equal prominence, but d2 should still increase by the same amount as d1 when p1 is raised. In other words, Ladd's model predicts that the function relating d1 to d2 should have a slope of 1. To examine this prediction, Figure 3 plots the results of the two tasks in terms of ERB-rate scale (E) distances from the actual baseline. It is evident that in both cases the slopes (solid lines) are considerably less than 1; in other words, d2 increases more slowly than d1. (The situation is similar in a Hz or st plot.) This aspect of the data seems strongly inconsistent with Ladd's model.

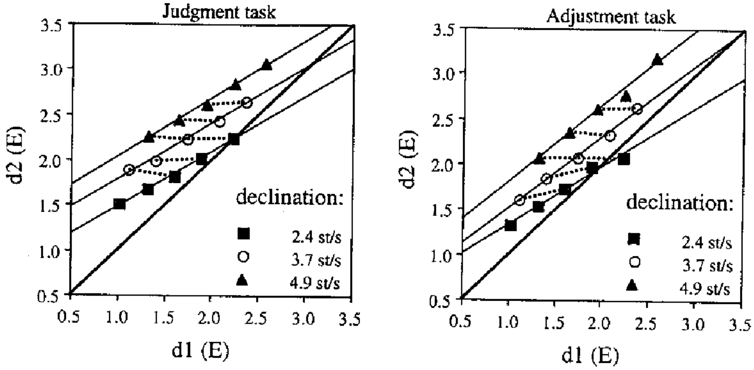


Figure 3. *Results of experiment 1, replotted in terms of distances (in E) from the declination line, with straight lines matched to the data points for each declination rate. The dotted lines connect points for the same p1 value.*

It is obvious from Figure 3 that p1 and p2 are generally not equidistant from the actual baseline when s1 and s2 are judged to be equally prominent, as already noted by Terken (1991). Terken's (1993) revised model includes the bottom of the pitch range as an additional parameter. Can it account for the present data? Table III lists values of p1 and p2, and of d1 and d2 in Hz, in E, and in st, for the judgment task. The last column lists the predictions of Terken's model, using the parameter values that fit the data of Terken (1991), viz. a = 0.9 and b = 0.23. These predictions should be compared to the d2 (Hz) column, and major discrepancies can be seen at the lower p1 values within each declination condition. Thus the present data cannot be fitted into the same parameters as the earlier results, which were obtained with reiterant male speech. For the present data, the average slope, the "a" coefficient in Terken's model, is 0.6 (the average value in Table II), which is considerably lower than Terken's earlier estimate of 0.9. This points to a problem with the model.

Table III. *Distances of p1 and p2 from the interpolated baseline in Hz, E, and st (judgment task data only), and d2 values (in Hz) predicted by Terken's (1993) model with preset parameters (experiment 1).*

Decl.	p1	d1	d1	d1	p2	d2	d2	d2	pred. d2
(st/s)	(Hz)	(Hz)	(E)	(st)	(Hz)	(Hz)	(E)	(st)	(Hz)
2.4	247	49	1.01	3.83	246	71	1.50	5.90	52
	263	65	1.32	4.91	255	80	1.67	6.52	66
	278	80	1.60	5.88	263	88	1.82	7.05	80
	294	96	1.89	6.84	273	98	2.01	7.70	94
	313	115	2.22	7.93	285	110	2.24	8.44	111
3.7	278	57	1.11	3.97	275	93	1.89	7.15	64
	294	73	1.40	4.94	280	98	1.98	7.46	79
	313	92	1.73	6.03	293	111	2.23	8.30	96
	333	112	2.07	7.10	305	123	2.43	8.94	114
	351	130	2.36	8.01	317	135	2.64	9.61	130
4.9	313	71	1.31	4.45	303	115	2.26	8.26	82
	333	91	1.64	5.53	313	125	2.44	8.83	100
	351	109	1.94	6.44	323	135	2.61	9.37	116
	370	128	2.24	7.35	337	149	2.83	10.10	133
	392	150	2.57	8.35	351	163	3.06	10.81	153

Terken's (1993) model was intended to capture the earlier finding (Terken, 1991) that narrowing the pitch range in the beginning of the utterance (i.e. reducing d1 by raising the baseline while keeping p1 constant) resulted in a narrowing of the pitch range (i.e. a reduction of d2) near the end of the utterance. To test his prediction, in Fig. 3, d1 values which represent identical p1 values under the different declination conditions are connected by dotted lines. As can be seen from the figure, in this data a given p1 value tended to be associated with a constant excursion size d2, regardless of the declination rate and thus regardless of d1. This means that narrowing of the pitch range did not play a role in this experiment. Moreover, the observation that a constant value of p1 is associated with a constant value of d2 indicates that the effect of decreasing d1, which would normally lead to a decrease in prominence, is cancelled by the effect of raising the onset of the baseline. This indicates that these two factors, i.e. the onset of the declination line and the excursion size of the first peak, do influence the prominence of s1 in opposite directions.

Before discussing this matter further, we will present the results of a second experiment that used male speech, which makes it more comparable to Terken's (1991) study. Only an adjustment task was used in this experiment, as the two tasks in experiment 1 had yielded rather similar results.

5.3 EXPERIMENT 2

5.3.1 Method

5.3.1.1 Materials. A male Dutch speaker spoke the same sentence as in experiment 1, "Amanda gaat naar Malta". The speech was digitized at 20 kHz and low-pass filtered at 10 kHz. The total utterance was about 1,410 ms in duration; the critical /mɑn/ and /mɑl/ syllables had durations of about 230 ms (/m/: 80 ms; /ɑn/: 150 ms) and 270 ms (/m/: 80 ms; /ɑl/: 190 ms) respectively. The utterance was LPC-coded (32 coefficients, 10 ms frames), and the pitch contour was determined. This contour was then subjected to stylization, which was straightforward as the unaccented syllables could easily be matched to a straight (i.e. exponentially declining) declination line. The declination rate was 3.9 st/s, which is almost exactly the rate predicted by the Dutch synthesis model (3.8 st/s). The two accent peaks (originally 183 and 139 Hz, respectively) were approximated by "pointed hat" patterns modeled on those of the stylized female source utterance of experiment 1; they were modified to be similarly aligned with the segmental structure. That is, the durations of rise, top, and fall portions were 110, 70, and 150 ms respectively for both peaks, with the rise starting 30 ms prior to vowel onset. The pitch of the top portion declined at the baseline rate, so that the maximum pitch occurred at the end of the rise. The temporal distance between p1 and p2 was 650 ms. The pitch contour of the stylized source utterance is shown by the solid line in Figure 4. There were three voiceless gaps: for /X/ and /t/ in "gaat", and for /t/ in "ta"; the last one again occurred during the fall of p2.

As in experiment 1, a set of 105 pitch contours was constructed by combining 3 declination rates, 5 values of p1, and 7 values of p2, selected from a range of 9 p1 and 15 p2 values. p1 values ranged from 150 to 238 Hz, and p2 values from 119 to 267 Hz, in steps of approximately 1 st. The three declination rates chosen were 2.4, 3.9 (original), and 5.5 st/s. Figure 4 presents a schematic illustration of these parameters. The precise p1 and p2 combinations are presented in Table IV.

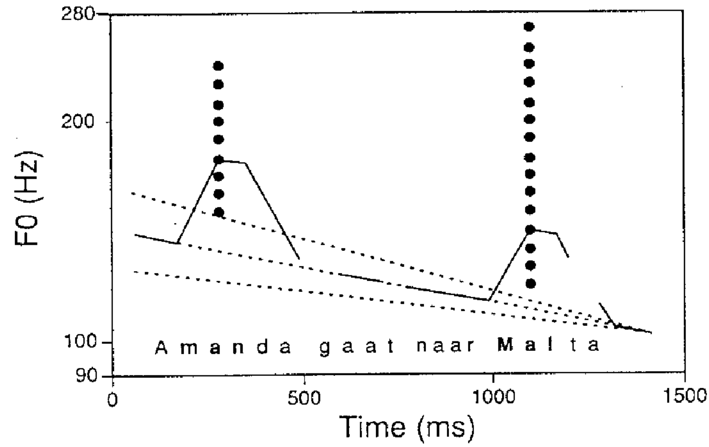


Figure 4. *Schematic illustration of stimulus parameters in experiment 2.*

5.3.1.2 Subjects and procedure. The subjects were 10 members of the research staff at IPO, 5 women and 5 men, all native speakers of Dutch, who volunteered to participate. None of them is involved in speech research, and none had participated in experiment 1. Instructions and testing procedure were identical to those in the adjustment task of experiment 1. Again, the data were averaged over adjustment runs in which p2 started at its highest and its lowest value in order to counter the effect of hysteresis.

Table IV. *Range of p2 values for each p1 value in experiment 2.*

Decl.	p1	p2 (Hz)														
(st/s)	(Hz)															
2.4	150	119	126	133	141	150	159	168								
	159		126	133	141	150	159	168	177							
	168			133	141	150	159	168	177	189						
	177				141	150	159	168	177	189	200					
	189					150	159	168	177	189	200	211				
3.9	168			133	141	150	159	168	177	189						
	177				141	150	159	168	177	189	200					
	189					150	159	168	177	189	200	211				
	200						159	168	177	189	200	211	225			
	211							168	177	189	200	211	225	238		
5.5	189					150	159	168	177	189	200	211				
	200						159	168	177	189	200	211	225			
	211							168	177	189	200	211	225	238		
	225								177	189	200	211	225	238	250	
	238									189	200	211	225	238	250	267

5.3.2 Results and discussion

Figure 5 presents the points of equal prominence. Again, the data points all fall below the diagonal, which replicates the basic p2 < p1 effect. The effect was present in eight of the ten subjects. Two subjects adjusted p2 to values higher than p1 when p1 was above 200 Hz.

The earlier finding that the p1-p2 difference increases with the value of p1 was less pronounced here. Table V shows the slopes of lines fitted to the individual subjects' data in the three declination conditions, as well as the averages of these slopes. For two subjects (the ones mentioned in the preceding paragraph) the slopes were greater than 1, so the average slopes were not significantly different from 1 [$F_{(1,9)} = 3.56$, $p < .10$]. Still, for the majority of the subjects the slopes were less than 1, and a combined analysis with the slopes of experiment 1b did not yield a significant difference between experiments [$F_{(1,15)}$ = 2.20, p < .16] while at the same time showing a significant overall difference from unity [$F_{(1,15)}$ = 17.01, p < .001]. The difference in slopes across declination conditions was not significant [$F_{(2,18)}$ = 1.30, p < .30].

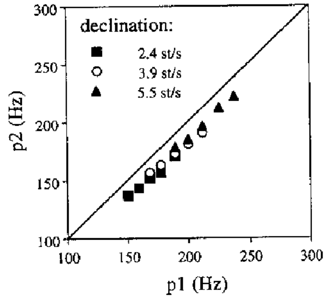


Figure 5. *Equal-prominence p2 values as a function of p1 value in experiment 2.*

As in experiment 1, there was a small effect of declination rate on the value of p2 matched to a given p1. However, it was only marginally significant in ANOVAs on the responses to p1 values shared by declination conditions [low vs. medium: $F_{(1,9)}$ = 4.86, p < .06; medium vs. high: $F_{(1,9)}$ = 7.77, p < .03; all three: $F_{(2,18)}$ = 3.10, p < .07].

Table V. *Slopes of straight lines describing the p1-p2 equal-prominence relationship in each declination condition in experiment 2.*

Subject	Declination rate		
	2.4 st/s	3.9 st/s	5.5 st/s
#1	0.76	0.36	1.05
#2	1.09	1.28	1.16
#3	1.11	1.67	1.16
#4	0.83	0.57	1.06
#5	0.86	0.71	0.84
#6	0.80	0.36	0.65
#7	0.75	0.74	0.70
#8	0.87	0.57	0.87
#9	0.81	0.91	1.02
#10	0.84	0.86	0.91
Average	0.87	0.80	0.94

The results are presented in terms of distances from the baseline in Table VI and in Figure 6. It can be seen that when the declination rate was low, equal prominence actually corresponded to equal distance from the baseline. At the higher declination rates, however, p2 was again higher above the baseline than p1 (d1 < d2). The last column of Table VI lists the predictions of Terken's (1993) model, with the parameters preset to the values that fitted to his earlier data for male reiterant speech (a = 0.9, b = 0.23). Again, the model seriously underestimates d2, except at the lowest rate of declination. The problem lies in the intercepts rather than the slopes of the d1-d2 functions (Fig. 6). The average slope was 0.87, which is very close to the value of 0.9 estimated from Terken's earlier data. But as can be seen in Table VI, from the discrepancies between the obtained and predicted values it becomes clear that even the present male speech data collected in an adjustment task cannot be fitted into the same model as the earlier reiterant speech data.

Table VI. *Distances of p1 and p2 from the interpolated baseline in Hz, E, and st, and d2 values (in Hz) predicted by Terken's (1993) model with preset parameters (experiment 2).*

Decl. (st/s)	p1 (Hz)	d1 (Hz)	d1 (E)	d1 (st)	p2 (Hz)	d2 (Hz)	d2 (E)	d2 (st)	pred. d2 (Hz)
2.4	150	30	0.77	3.90	137	30	0.79	4.25	31
	159	39	0.99	4.90	143	36	0.93	4.95	39
	168	48	1.20	5.80	152	45	1.17	6.06	47
	177	57	1.41	6.70	157	50	1.28	6.57	55
	189	69	1.68	7.90	171	64	1.62	8.10	66
3.9	168	35	0.86	4.00	157	47	1.19	6.08	39
	177	44	1.07	4.90	163	53	1.35	6.80	47
	189	56	1.34	6.10	173	63	1.58	7.79	58
	200	67	1.58	7.10	181	71	1.77	8.59	67
	211	78	1.82	8.00	191	81	1.99	9.49	77
5.5	189	41	0.96	4.20	178	64	1.59	7.67	47
	200	52	1.20	5.20	185	71	1.74	8.30	57
	211	63	1.44	6.10	196	82	1.99	9.34	67
	225	78	1.73	7.30	212	98	2.32	10.66	81
	238	91	1.99	8.20	222	108	2.55	11.53	92

On the other hand, the predictions of Ladd's (1993) model fit to these data better than those of experiment 1. Rate of declination had only a small effect on the relation between p1 and p2 (Fig. 5), and the functions relating d1 to d2 have a slope close to 1 (Fig. 6).

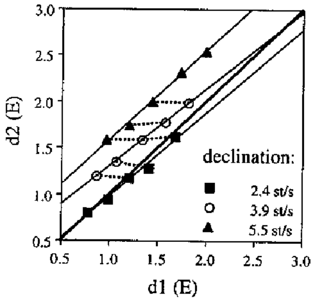


Figure 6. *Results of experiment 2, replotted in terms of distances (in E) from the declination line, with straight lines matched to the data points for each declination rate. The dotted lines connect points for the same p1 value.*

In this figure, as in Fig. 3, d1 values going with a given p1 value under the different declination conditions are connected by dotted lines. It can be seen that there is, again, a strong tendency for a given p1 to be associated with a constant d2, regardless of the declination rate and thus regardless of d1.

5.4 GENERAL DISCUSSION AND CONCLUSIONS

The present study extends previous findings on the perception of the relative prominence of two pitch accents within a single utterance, and it shows that the situation is somewhat more complex than predicted by the models envisioned so far. By varying declination rate and the value of p1 in a semi-independent fashion, the effects of these two factors on prominence perception could be assessed. The effect of declination rate was surprisingly small in terms of peak maxima, but it was large in terms of excursion sizes from the actual baseline.

From the results of experiments 1a and 1b it can be concluded that prominence judgments and adjustments give basically the same results. The adjustment method has two important advantages over the judgment method.

First, the points of subjective equality are obtained directly from the (average) adjusted values, without the necessity of calculating psychophysical curves. Secondly, the number of stimuli which the subjects will choose from the continuum to listen to is smaller for the adjustment experiment, since the stimuli which are most unlikely to represent equal prominence will be skipped by them. As a consequence, the length of an adjustment experiment was only half that of the judgment experiment.

The results of experiments 1 and 2 are sufficiently consistent to support the conclusion that prominence relationships in the female and male pitch ranges are perceived similarly. The pitch peak maximum p2 is usually lower than p1 when s2 and s1 are judged to be equally prominent. This robust effect has been observed in all previous studies. It is commonly attributed to a perceptual compensation for expected topline declination. That is, the explanation appeals to a fact about natural speech that has been internalized by listeners (Pierrehumbert, 1979).

Terken (1991) also observed that the difference between p1 and p2 tends to increase as the value of p1 increases, but then it reflected a joint effect of p1 value and declination rate. These results suggest that the increase depends mainly on p1, and that declination has a small effect which is nevertheless significant. Although the prominence-matched value of p2 is smaller than that of p1, distance d2 is generally larger than d1, whether they are measured in Hz, E, or st, which is not consistent with a model according to which equal prominence means equal distance from the baseline (cf. Terken, 1991).

Ladd's (1993) proposal of a "virtual" reference line circumvents the problem of trying to account for prominence perception in terms of physically measurable parameters, but it has its own problems, as pointed out by Terken (1993). Although Ladd's predictions were consistent with the outcomes of experiment 2, the small but consistent effect of declination on the p1-p2 relationship and the unequal change in d1 and d2 with changes in p1 in experiment 1 are inconsistent with his model.

The finding that, as p1 increases, the difference between p1 and p2 must increase in order to maintain equal prominence also appears to be consistent with the findings of Gussenhoven and Rietveld (1988). From their data it may be derived that, with declination kept constant and an approximately equal p1-

p2 difference, the prominence of the p2 syllable increases relative to the p1 syllable as p1 gets higher: for small values of p1, the p1-p2 difference employed by Gussenhoven and Rietveld made their listeners judge p2 to be less prominent than p1; for high p1, the same p1-p2 difference was judged to give approximately equal prominence to p1 and p2.

The results of experiments 1 and 2 were also consistent in showing a strong tendency for d2 to be constant for a given p1 under the three declination conditions (and thus regardless of the size of d1). It is important to notice that the "true" declination baseline was anchored to the end of the utterance, so that changes in the rate of declination mainly affected the pitch at the start of the utterance and much less at the end of it. The finding that d2 is constant for a given value of p1 might indicate that the height of the onset as well as the excursion size of the pitch peak on s1 influence its prominence in opposite ways. If d1 increases due to a decrease of the declination rate, s1 would be perceived as being more prominent, but this effect is then cancelled by the lowering of the onset of the pitch contour.

A problem with this interpretation is that it is not clear why on the one hand the pitch peak maximum p1 determines the prominence of s1, while on the other hand only a distance measure, d2, would determine the prominence of s2. One alternative possibility is that listeners have some other reference than the baseline for judging the relative height of maximum p1: instead of making use of the information about the lowest pitch level in the speaker's range which is normally only available at the end of an utterance, they might use other information which is locally available when building a reference framework. One plausible candidate seems to be the spectral build-up of the speech signal, providing information about the speaker's performance when producing a certain pitch-peak maximum, his voice characteristics giving clues about his pitch range. This would also be consistent with the findings by Rump and Hermes (in press, see also chapter 4), who found that variation of the height of the actual baseline did not influence the adjusted heights of pitch levels on the toplines of their stimuli when listeners adjusted relative prominence to be equal. It would also be consistent with the findings by Gussenhoven and Rietveld (1993). They hypothesized that when judging the relative height of the first pitch peak, listeners need information about preceding low-pitch levels spanning at least two unaccented syllables. In our experiments, there was only one unaccented syllable. One may guess that the results of these experiments might have been quite different if stimuli with relativly long unaccented onsets would have been used.

The assumption of an internal reference for judging the maximum of a pitch peak early in an utterance, which is based on information locally available in the speech signal, can circumvent the problem of how listeners create a low reference line, which is based on information which is in general available only at the end of an utterance. This proposal of an internal reference is not inconsistent with the models by Terken (1993) and Ladd (1993). All assume that listeners make use of their knowledge about a speaker's pitch range which is assumed to decline towards the end of an utterance. The main difference between these models and our explanation of the data is that, in our view, listeners make use of information which is available ''on-line'' in order to build a reference framework, whereas in the other models listeners need information about the end of the utterance before a framework can be built.

In summary, the results of these experiments, involving both a judgment and an adjustment task as well as a comparison between perceived prominence in male and female pitch ranges, are highly consistent. In addition, they are consistent with most of the earlier findings. Firstly, the value of p2 is generally smaller than that of p1 when the pitch-accented syllables are perceived as being equally prominent. Secondly, the p1-p2 difference increases with higher values of p1. Thirdly, the effect of declination rate, although it is small, is generally significant. It is, however, not completely clear how this latter finding should be interpreted. At higher declination rates, a given excursion size d1, implying a higher value of p1, causes the first pitch-accented syllable to be perceived as being more prominent, which is reflected by an increase of both p2 and d2. However at higher declination rates, d2 turns out to be constant for a given pitch-peak maximum p1, although excursion size d1 decreases. A constant value of d2 implies that the prominence of s2 is constant. On the basis of the finding that d2 is constant for a given value of p1, it has to be assumed that the value of the pitch peak maximum p1 determines the prominence of s1. This would mean that not only information about pitch but also information about the speaker's voice characteristics might help the listener when creating a reference frame for judging the prominence lent by the first pitch peak. Together with peak-to-peak scaling, this would ultimately imply that no baseline, neither actual nor virtual, is needed for judging the prominence lent by this pitch peak, but that instead the pitch maximum is used together with voice-source information.

In conclusion, the maximum of a pitch peak seems to be scaled relative to the low pitch levels in the preceding unaccented syllables. However, in the case that there is only one preceding unaccented syllable, containing only little information about the preceding low pitch level, the maximum of a peak seems to be scaled within a speaker's pitch range on the basis of voice characteristics.

6 FOCUS CONDITIONS AND THE PROMINENCE OF PITCH-ACCENTED SYLLABLES[16]

ABSTRACT

The purpose of the present study is to find out how the pitch peak heights on two pitch-accented syllables in one utterance relate to different focus conditions. The focus conditions are neutral focus, double contrastive focus, and single contrastive focus on either the first or the second pitch-accented syllable. In experiment 1, subjects adjusted the height of one of two pitch peaks, so as to make the pitch contour express different focus structures. No systematic relationship was found between different fixed heights of one peak and the adjusted heights of the other, which suggests the existence of target values for focus-related pitch peaks. In experiment 2, listeners judged which of the four focus structures was most likely represented by a given relation between peak heights. The results show that some pitch contours are ambiguous with respect to focus, but the majority of them is classified unanimously as signalling only one possible focus structure. The present results also shed new light on some unexplained findings of earlier prominence experiments.

6.1 INTRODUCTION

A number of earlier experiments on prominence perception have compared the relative prominences of two accented syllables, with rising-falling pitch movements, in one utterance. (Pierrehumbert, 1979; Rietveld and Gussenhoven, 1985; Gussenhoven and Rietveld, 1988; Terken, 1991, 1994;

[16] Published as H.H. Rump and R. Collier: "Focus conditions and the prominence of pitch-accented syllables," Language and Speech 39 (1996), 1-17. Parts of the research reported here were presented at the XIIIth International Congress of Phonetic Sciences (ICPhS 95), Stockholm, August 1995, as H.H. Rump: "Influence of focus structures on tonal targets of pitch peaks," Proceedings Vol. 3, pp. 664-667. The data of experiment 2 were also published as H.H. Rump and R. Collier: "Pitch-peak height and focus," in the IPO Annual Progress Report 30 (1995), 45-50.

Repp *et al.*, 1993; Ladd *et al.*, 1994). These experiments involved both prominence judgments and prominence adjustments: in the former listeners had to indicate which of two pitch-accented syllables was the more prominent, while in the latter they varied the height of the pitch peak on the second pitch accent in order to make its prominence equal to the prominence of the first. Repp *et al.* (1993) demonstrated that the results for both types of experiments did not differ significantly. The first observation to be made about the results of these experiments is that, in the case of equal prominence, the value of 'p2' (the absolute height of the peak on the second pitch-accented syllable 's2') was in general somewhat smaller than that of 'p1' (the absolute height of the peak on the first pitch-accented syllable 's1'). This was found to hold for both nonsense and meaningful utterances, and for utterances with and without baseline declination (the downtrend of pitch normally observed across declarative utterances). Terken (1994) found that the effect held to the same extent for a pitch peak which is not the last peak in the utterance, which led him to conclude that it was not induced by expectations about ''final lowering'', i.e. the extra lowering of a peak on the final pitch accent which is often observed in production (Liberman and Pierrehumbert, 1984). Rather, the observation that the value of p2 is smaller than that of p1 when s1 and s2 are equally prominent should be attributed to perceived or expected declination.

The second observation in the earlier experiments was that, if s1 and s2 are equally prominent, the values of p1 and p2 are related to each other in a rather simple way. If the value of p1 increases, that of p2 increases, although the rate of increase is somewhat smaller for p2 than for p1 (Terken, 1991, 1994; Repp *et al.*, 1993). Finally, it was observed by Pierrehumbert (1979) and Terken (1991) that for small pitch peaks on s1 the values of p2 tended to be larger than those of p1 when the pitch accents were perceived as equally prominent. Several explanations for these observations have been given. They are included as factors in the prominence model by Liberman and Pierrehumbert (1984, 191-192), which has been elaborated by Terken (1993). A last factor, which has not been included in the model but which is known to influence pitch peak heights, is the focus structure of the utterance. In previous experiments, focus conditions were not an independent variable. One goal of the present paper is to test a hypothesis concerning the possible influence of the focus structure of the test utterance on the results of previous prominence experiments.

The main goal of the experiments presented here is to find out the relationship between the relative heights of two pitch peaks and the focus structure of the utterance. It is already known from production experiments that different focus structures are mainly signalled by different heights of the pitch peaks, at least for rising-falling pitch movements (O'Shaughnessy, 1979; Liberman and Pierrehumbert, 1984; Eady and Cooper, 1986; Eady *et al.*, 1986; Horne, 1988; Pierrehumbert and Hirschberg, 1990; Välimaa-Blum, 1993; Ladd and Terken, 1995). The focus structure of an utterance helps the listener to detect which parts of an utterance are most informative according to the speaker. The speaker can mark these important parts either prosodically, or by changing the word order so that the focused part comes first. In the present paper, we will discuss only prosodic marking of focus by means of pitch. For example consider the utterance 'Amanda's going to Malta' in which the syllables /mɑn/ and /mɑl/ are the accented ones, s1 and s2, respectively. Four different focus structures may be elicited by asking the following questions:

- What is happening? (neutral focus)
- Is John going to Cyprus? No,... (double contrastive focus, on Amanda and on Malta)
- Is John going to Malta? No,... (single contrastive focus on Amanda)
- Is Amanda going to Cyprus? No,... (single contrastive focus on Malta).

The different answers to the questions are textually identical, but are assumed to be marked by different pitch contours. In general, neutral focus is known to be signalled by relatively low pitch peaks, while double focus is prosodically marked by pitch peaks which are rather high in the speaker's range (Bartels and Kingston, 1994). Under the single-focus conditions only one of the pitch-accented syllables is in focus while the other is defocused. Defocused syllables are the ones which would get a pitch peak if the utterance were spoken in a neutral declarative way. If defocused syllables are deaccented (i.e. having no pitch peak at all), they will still be marked by a longer duration and/or a higher amplitude so that they are considered to be stressed (see Sluijter, 1995, and the references therein).

These observations have been incorporated in a number of models that assume the existence of pitch-targets, in particular the "grid" model developed by

Bruce (1977) for Swedish and applied to Swedish by Gårding (1981) and to English by Horne (1988). In these models the grid represents the non-emphatic pitch register of a speaker. The heights of pitch peaks are scaled relative to the grid. These relative heights in the grid represent so-called target pitch levels. The target levels are assumed to be present in a speaker's mind and he or she produces these pitch levels in realising a pitch contour conveying a certain focus structure. The models were designed to be used in speech synthesis but, to our knowledge, have not been tested in formal perception experiments.

According to the grid models, pitch peaks should reach fixed target levels which differ for various focus conditions (see, e.g. Horne, 1988). According to the results from the prominence experiments, however, the heights of pitch peaks, and therefore the prominence of pitch-accented syllables, is not fixed but may vary gradually. This difference between the grid models and the observations in the prominence experiments may be attributed to the fact that they address the perceptual phenomena at different levels of abstraction. The prominence experiments are 'psychophonetic' experiments, in which listeners pay attention to relatively small details of the speech signal and are able to make gradual distinctions. The grid models speak to the phonology of pitch accents, and specify the *functional* relationship between pitch phenomena, which are more of an all-or-none nature (Bolinger, 1961).

In summary, the question we want to address here is how the prominence of two pitch-accented syllables is related to the focus structure of the utterance. Under the neutral-focus condition, none of the pitch accents in the utterance gets special attention. Under the double-focus condition, the two pitch accents are in (contrastive) focus at the same time. Under the single-focus conditions only one pitch-accented syllable is in contrastive focus while the other syllable is defocused.

In addition, we want to relate the outcome of the present experiments to the outcome of previous prominence experiments. There is evidence that different focus structures are prosodically signalled by different peak heights. In the previous prominence experiments prominence was varied by changing the heights of the pitch peaks. The underlying focus structure of the utterance may thus have been affected as well. We assume that "equal prominence" can occur under neutral-focus and double-focus conditions, and that a single-focus condition requires explicit ''unequal prominence'' of the pitch-accented syllables.

In experiment 1, the influence of focus on the *adjusted* heights of pitch peaks was tested. In the first part of experiment 1, listeners adjusted the value p2, the absolute height of the pitch peak on s2, in the second part, they adjusted the value of p1. In experiment 2, the influence of peak heights on the perceived focus condition was tested. Listeners decided which focus structures were signalled by the various pitch contours.

6.2 EXPERIMENT 1: PEAK-HEIGHT ADJUSTMENTS

In the earlier prominence experiments, subjects adjusted the heights of either the second or the last pitch peak in the utterance, but they never adjusted the first one. In the present experiment the subjects adjusted the second or the first pitch peak.

6.2.1 Method

The utterance used in the present experiments was the Dutch utterance *A'manda gaat naar 'Malta* (*A'manda is going to 'Malta*), which was produced by a male speaker. It contained two pitch-accented syllables s1 and s2, /mɑn/ and /mɑl/, with the associated peak values p1 and p2, respectively. The pitch accents in the original utterance had rising-falling pitch contours while the pitch in the unaccented syllables was determined by a declining baseline. The pitch contour was stylized according to the rules of the speech-synthesis system for Dutch, in use at the Institute for Perception Research, using straight lines in the ERB frequency domain (Patterson, 1976; Glasberg and Moore, 1990; Hermes and Van Gestel, 1991). The stylized rises and falls always had durations of 120 ms. After the rise, which started 50 ms before the onset of the vowel of the accented syllable, the pitch stayed on the topline for 30 ms. The original and the stylized pitch contours are shown in Fig. 1. The starting frequency of the baseline was 4.36 ERB-units E (137 Hz), the end frequency was 3.37 E (100 Hz). The duration of the utterance was 1.45 s. The rate of declination was about 0.7 E/s. The formula we used for translating Hz values into E values adapted from Glasberg and Moore (1990, p. 114):

$$E = 21.4 \cdot LN(0.00437 \cdot F + 1)/(LN10) \text{ (F in Hz).}$$

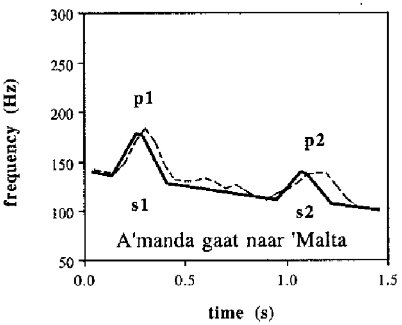


Figure 1. *Original (dashed line) and stylized (solid line) pitch contours of the utterance A'manda gaat naar 'Malta. The horizontal axis is the time axis (s), the vertical axis is the frequency axis (Hz). Excursion sizes of the peaks are measured from points on the baseline right below the tops of the peaks. For technical reasons the lines are drawn as straight lines in the linear frequency domain (Hz). In the stimuli, however, the pitch contours consisted of straight lines in the ERB-rate frequency domain (E). The pitch-accented syllables /mɑn/ and /mɑl/ are referred to as s1 and s2, respectively. The associated pitch-peak values are referred to as p1 and p2.*

Two kinds of adjustments were performed. In the first part of experiment 1, subjects adjusted the value of p2 while that of p1 was fixed. In the second part, subjects adjusted the value of p1 while that of p2 was fixed. Adjustments amounted to selecting the appropriate pitch contour from a prepared set of stimuli (see below). Pitch manipulations were performed using the PSOLA method (Hamon, Moulines, and Charpentier, 1989).

The task of the subjects was to adjust the height of a given pitch peak so that the resulting pitch contour would optimally realize a given focus condition. Different focus structures were suggested by making the test sentence sound as an answer to one of the four questions, which were already mentioned Introduction section:

- What is happening? (neutral focus)
- Is John going to Cyprus? No,... (double focus, on s1 and s2)
- Is John going to Malta? No,... (single focus on s1)
- Is Amanda going to Cyprus? No,... (single focus on s2).

Neutral focus was meant to give a non-contrastive reading of the utterance. Under the single-focus conditions only one of the pitch-accented syllables was in focus while the other was defocused. Under the double-focus condition, both s1 and s2 were in contrastive focus at the same time. The answer was always textually the same, i.e. Amanda's going to Malta, only its pitch contour was variable. The Dutch versions of the questions were printed on the computer screen while the Dutch test utterance '*Amanda gaat naar Malta*' was made audible through headphones.

6.2.1.1 Adjustments of p2. The first group of subjects adjusted the value of p2, i.e., the height of the pitch peak on s2, so that the utterance with the resulting pitch contour would be an appropriate answer to the question written on the screen. During each trial, the value of p1, the height of the pitch peak on s1, was fixed at one of three different values: 165, 183, or 202 Hz. Adjustments started at both extremes of the peak height continuum, the value of p2 ranging from 110 to 214 Hz, which corresponded to excursion sizes of the pitch peak of zero to 2.5 E above the baseline (measured perpendicular to the time-axis). The range was divided into nine steps of 0.25 E (about 1.5 semitones). Each adjustment was repeated twice, so that a subject completed four trials per question and per value of p1. The total number of adjustments was 48. The experiment was preceded by a short introduction in which the subject made six trial adjustments. The order of presentation was completely random.

Ten subjects participated. They were students and research staff of the institute. They were all native speakers of Dutch, and they were not phonetically trained.

6.2.1.2 Adjustments of p1. The same utterance was tested with a different group of ten phonetically naive subjects. None of them had participated in the first part of experiment 1. They adjusted the value of p1, the height of the pitch peak on s1, while p2 had one of three fixed values: 143, 160, and 179 Hz. The continuum of p1 values ranged from 131 to 267 Hz in eleven steps which were equidistant in E, viz. 0.25 E or about 1.5 semitones. The questions were the same as in experiment 1. The order of presentation was again random.

6.2.2 Results and discussion

6.2.2.1 Adjustments of p2. In order to counter the effect of hysteresis, the results were averaged across starting points. The average p2 heights were found to be 113, 161, 172 and 202 Hz for s1 in single focus, neutral focus, double focus, and s2 in single focus, respectively. A repeated-measures analysis of variance was performed with p1 height and focus condition as fixed factors and with subjects as replication factor. It revealed that the main effect of focus conditions on the adjusted values of p2 was highly significant ($F(3,27) = 70.8$, $p < 0.001$). The main effect of p1 on the adjusted value of p2 was non-significant ($F(2,18) = 2.94$, $p < 0.08$), nor was the interaction between focus conditions and p1 ($F(6,54) = 1.78$, $p < 0.12$). This indicates that for none of the focus conditions did the set values of p1 have a systematic effect on the adjusted values of p2. The results for the individual subjects, listed in Table A in the Appendix to this chapter for each of the focus conditions, are therefore averaged across the three different values of p1. It can be seen that every subject adjusted the value of p2 to be almost minimal, i.e. having zero excursion size, when s1 was in single focus, and almost maximal when s2 was in single focus. In the case of neutral and double focus, there was more variation between the subjects' behaviour. Some of them adjusted p2 to somewhat higher values under the double-focus than under the neutral-focus conditions, whereas others did the reverse.

In Table I the results, averaged across subjects, are presented for the three values of p1. A planned comparison showed that the difference between the adjusted peak heights under the neutral-focus and double-focus conditions was not significant ($F(1,27) = 0.69$, $p > 0.05$), although the values of p2 tended to be larger under the double-focus than under the neutral-focus condition.

For the single-focus conditions the results were as expected: if s2 was in single focus, the value of p2 was adjusted almost as large as possible, and if s1 was in single focus, the value of p2 was adjusted to be as small as possible. Unexpectedly, however, it was found that the adjusted value of p2 did not differ significantly between the neutral-focus and double-focus conditions. This issue will be further discussed below.

Table I. *Adjusted values of p2 (Hz) under four different focus conditions and for the three fixed values of p1 (Hz), averaged across subjects.*

	intended focus conditions			
	s2	neutral	double	s1
p1				
165	201	157	170	116
183	199	162	167	111
202	206	164	178	111

Another striking finding was that there was no systematic influence of p1 height on the adjusted height of p2, contrary to what was expected on the basis of the earlier prominence experiments. The fact that this did not occur supports the assumption that language users may be aware of target values for the absolute peak heights under the different focus conditions.

As mentioned above, an important finding was that listeners did not seem to discriminate between neutral focus and double focus on the basis of the value of p2, although the grid models as well as emphasis-production experiments (e.g. Ladd and Terken, 1995) strongly indicate that the pitch range of the utterance should be expanded for the double-focus structure. Therefore the question is: on what other basis do listeners decide whether the pitch range is expanded or not? A possible cue for perceiving local pitch ranges may then be the height of the first peak. This would imply that if the listeners were allowed to adjust p1, they would choose a larger value under the double-focus than under the neutral-focus condition. This was tested in part 2 of experiment 1.

6.2.2.2 Adjustments of p1. On the basis of the results of the p2 adjustments, it was expected that the task might be difficult for the condition under which s1 was in single focus, since in that case the excursion size of the pitch peak on s2 should be zero. The subjects were therefore told in advance that the pitch contours might not sound optimal in every instance but that they should try to match the contour as closely as possible.

Again, the results were pooled across starting points. The average adjusted p1 heights turned out to be 147, 175, 212, and 249 Hz for s2 in single focus, neutral focus, double focus, and s1 in single focus, respectively. As in part 1, a repeated-measures analysis of variance with p2 height and focus condition as fixed factors and with subjects as replication factor showed that the effect of focus conditions on the adjusted values of p1 was highly significant ($F(3,27) = 58.0$, $p < 0.001$). The main effect of p2 on the adjusted value was non-significant ($F(2,18) = 1.27$, $p < 0.31$). As in part 1, the interaction of focus conditions and p2 turned out to be non-significant ($F(6,54) = 1.23$, $p < 0.31$). The value of the fixed pitch peak, therefore, did not systematically influence the adjusted heights of the other.

The results for the individual subjects, listed in Table B in the Appendix to this chapter for each of the focus conditions, were again pooled across the three heights of the fixed peak on s2. It can be seen that the results showed the same trend for each of the subjects: p1 was set to a high value if s1 was in single focus and it was set to a low value if s1 was defocused. In addition, it can be seen that the values of p1 were larger under the double-focus than under the neutral-focus condition for every single subject. A planned comparison showed that this difference was significant ($F(1,27) = 5.03$, $p < 0.05$). The results averaged across the subjects are listed in Table II.

Table II. *Adjusted values of p1 (Hz) under four different focus conditions and for the three fixed values of p2 (Hz), averaged across subjects.*

	intended focus conditions			
	s2	neutral	double	s1
p2				
143	149	175	209	250
160	141	175	209	250
179	150	174	219	249

As in the first part of experiment 1, single focus on the pitch-accented syllable with the adjustable peak height, here s1, was signalled by a large height of the focal peak. In contrast to the findings in part 1, however, the excursion size of the peak on s1 was significantly larger than zero when s1 was defocused, viz. the excursion size of the pitch peak on s1 was about 0.4 E (about 16 Hz or 2 semitones). These findings suggest that the excursion size of the pitch peak on the pre-nuclear accent should indeed be larger than zero although s1 is defocused. This is sometimes called a rhythmical accent.

It was found that the values of p1 are adjusted to be larger under the double-focus condition than under the neutral-focus condition. As expected, under the neutral-focus condition, the adjusted value of p1 was slightly higher than the average value of p2. As a result, the excursion sizes of the pitch peaks on s1 and s2 were about equal. Under the double-focus condition, however, the adjusted value of p1 was much larger than the average value of p2, so that the excursion size of the pitch peak on s1 was much larger than that on s2. This holds for the performance of almost every single subject.

From experiment 1, it has become clear that the focus condition is crucial for the adjusted peak heights. The results for the single-focus conditions show that for both the adjustments of p1 and p2 syllables in single focus require large pitch peaks, defocused syllables require secondary accents or no accents at all, depending on their position relative to the nuclear accent. Defocused pre-nuclear pitch peaks should have an excursion size larger than zero, up to about 2 semitones, but defocused post-nuclear syllables should be deaccented (i.e., there should be no pitch peak at all).

The results for the neutral-focus and the double-focus conditions were not as expected, viz. that the values of p1 and p2 would be larger under the double-focus than under the neutral-focus condition. It was found that double focus was prosodically marked only by significantly larger p1 values, whereas the values of p2 were about the same under these two conditions.

Another interesting finding from experiment 1 is that the exact height of the fixed peak did not systematically influence the adjusted height of the variable one. This may indicate that the listeners have disregarded the exact height of the fixed peak and have chosen a target value instead. As a result, only the adjustable peak heights were found to determine the differences between pitch contours. The assumption of internalized target values might also explain why the subjects did not find it difficult to perform the task. Listeners seem to have very clear ideas about how the pitch contour should sound under different focus conditions. In addition, the high inter-subject agreement (see Tables A and B in the Appendix to this chapter) for most of the focus conditions indicates that different listeners have more or less the same target values in mind, which may thus be part of the language phonology.
Ladd and Terken (1995) also found evidence for the existence of target values for pitch peaks: in an emphasis-production experiment, high and low values in the various pitch contours turned out to be relatively constant, not only for a given speaker but also for a given emphasis instruction.

If the average p1 and p2 values, found in our adjustment experiment, represent the 'true' target values for the different focus conditions, then for each of the four focus conditions the mean p2 value (from part 1 of experiment 1) and the mean p1 value (from part 2) can be combined into a single 'prototypical' pitch contour. In Fig. 2, the resulting pitch contours for each of the focus conditions are displayed.

These prototypes have rather simple characteristics, which would make it very easy for a listener to recognize any of the four focus condition. For example, single focus on s1 would then be signalled by the absence of a pitch peak on s2, whereas single focus on s2 is obtained if the value p2 is large and the value of p1 is small. The neutral-focus condition would then be marked by the fact that the baseline and the topline, connecting the first and second peaks, are about parallel, implying equal excursion sizes of the first and second peaks. Double focus would be marked by the fact that the value of p1 is much larger than that of p2, so that the topline is much steeper than the baseline. We were wondering whether we would find these target values in an experiment in which listeners could adjust the values of p1 and p2 at the same time. Unfortunately, however, this task turned out to be much too difficult to perform.

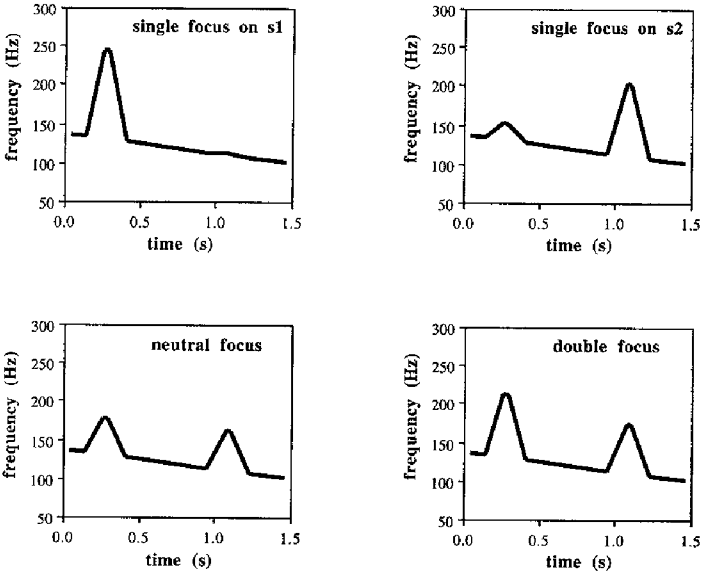


Figure 2. *Prototypical pitch contours for each of the different focus conditions, obtained from combining the adjusted values of p1 and p2, i.e. the averages displayed in Tables A and B (Appendix).*

Informal listening tests with resynthesized versions of the four prototypical pitch contours showed that the two single-focus contours were most likely to be perfectly recognizable as such. The neutral-focus and double-focus contours, however, seemed to be much more difficult to discriminate. It is therefore not clear whether listeners are indeed able to recognize an intended focus condition when they hear a prototypical or some other pitch contour. This is tested in experiment 2 in which subjects performed a judgment task instead of an adjustment task, which turned out to be relatively easy to perform.

6.3 EXPERIMENT 2: FOCUS JUDGMENTS

From the results of experiment 1 it may be concluded that a rather small difference between otherwise identical pitch contours, for example a slightly larger difference between the values of p1 and p2 under the double-focus than under the neutral-focus condition, could already present a cue to the listeners about the intended focus. Because the perceptual difference between the prototypical pitch contours for neutral and double focus appeared to be rather small, it is tested in a formal perception experiment whether listeners can recognize focus structures on the basis of the pitch contour alone.

6.3.1 Method

6.3.1.1 Stimuli. The same utterance as in experiment 1 was used, i.e. *A'manda gaat naar 'Malta*. The pitch contour was stylized as before. The excursion sizes of the first and second pitch peaks ranged from zero to 3 E above the baseline, spanning excursion sizes up to about 12 semitones or one octave. The ranges were each divided into 6 steps which were equidistant in E. The values of p1 ranged from 131 (zero excursion) up to 267 Hz, for p2 they ranged from 110 to 238 Hz. The combination of seven p1 values and seven p2 values resulted in forty-nine different pitch contours. The contour with both pitch peaks having zero excursions was not presented, so that the subjects judged 48 different pitch contours. Pitch manipulations were performed using the PSOLA method.

6.3.1.2 Procedure. Subjects were tested individually. They were seated behind a computer terminal. They listened through headphones to utterances resynthesized with each of the pitch contours described above. The subjects

were shown four questions on the screen and they had to decide to which of them the utterance was the most likely answer. They could listen to the utterance as frequently as they liked. After making a choice, the subjects pressed the corresponding key. The answer was recorded automatically, and the next stimulus was given. The subjects were thus forced to make a choice in every instance. Before the test started, twelve trial judgments were made. Within the same session, the total set of (12 plus 48) stimuli was presented twice, so that the total number of judgements per subject was 120.
Ten subjects participated in this experiment. They were native speakers of Dutch, having had no special training or education in phonetics. None of them had taken part in experiment 1.

6.3.2 Results

The average consistency within subjects, defined as the proportion of judgments which were identical across the first and the second presentations, turned out to be 74.7 percent. It means that, on average, 36 of the 48 pitch contours led to the same response in the first and second presentations. The between-subjects agreement for each of the focus conditions is graphically represented in a so-called 'density plot' (Fig. 3).

Figure 3 shows the percentage of judgements for each of the different focus structures under the different s1-s2 excursion sizes. Along the horizontal and vertical axes, the excursion sizes of the first and second pitch peaks are indicated, respectively. The total number of judgments for each p1-p2 combination was 20 (ten subjects judged each pitch contour twice). Since the experiment had a four-interval-forced-choice design, the a priori probability of a random score was 25 percent. The four areas are labelled according to the perceived focus structures and indicated by different shading. Each area is limited by a solid line labelled '50'. Within that boundary the percentage of judgments for a certain focus structure is greater than 50. Within each area, the 75-percent boundaries, labelled '75', are indicated. Boundaries were calculated using the Matlab statistics program. The dashed line indicates equal values of p1 and p2. The closed circles represent the prototypes which were hypothesized from the results of experiment 1. The open circles indicate p1-p2 combinations judged "equal" in prominence experiments (data from Gussenhoven *et al.*, to appear).

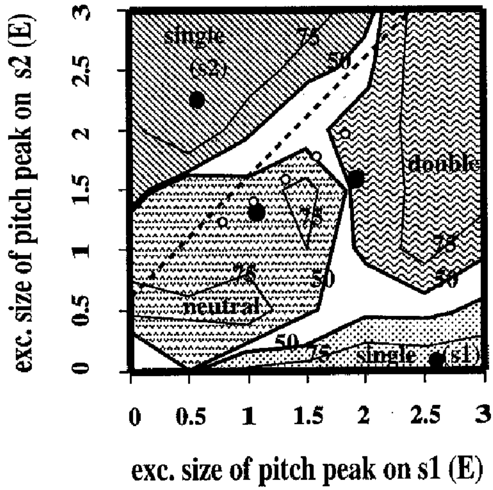


Figure 3. *Results of experiment 2. Density plot of the judgments for each of the excursion-size combinations. Along the axes, the excursion sizes of the pitch peaks on s1 and s2 are displayed. Zero excursion sizes correspond to values of 131 Hz (p1) and 110 Hz (p2). Largest excursion sizes, being 3 E, correspond to p1 and p2 values of 263 and 238 Hz, respectively. The a priori probability of obtaining any score is 25 percent. The shaded areas correspond to scores larger than 50 percent. The dashed line represents equal values of p1 and p2. Closed circles indicate the p1-p2 combinations which made up the prototypes hypothesized on the basis of the findings in experiment 1. Open circles represent the data of a prominence experiment (Gussenhoven et al., submitted).*

It can be seen that combinations of small values of p1 and p2 resulted in neutral-focus judgments and that combinations of large values resulted in double-focus judgments. In general, the listeners perceived s2 to be in single focus when the value of p2 was larger than that of p1. They perceived s1 to be in single focus only when the second pitch peak was almost absent.

6.3.3 Discussion

The high within-subject consistency (almost 75 percent) indicates that subjects have clear ideas about the characteristics a pitch contour should have in order to signal a certain focus condition. The large between-subjects correspondences which can be inferred from Fig. 3 indicate that subjects to a large extent share the same criteria. In addition, it can be seen that there is a clear asymmetry between the two single-focus areas. The area in the upper left-hand corner, indicating combinations of p1 and p2 that signal single focus on s2, is much larger than the one in the lower right-hand corner which covers the combinations of p1 and p2 that signal single focus on s1. This is comparable to the finding in experiment 1 that with single focus on s1, post-nuclear s2 should preferably have no pitch peak, whereas with single focus on s2 the pitch peak on pre-nuclear s1 may stand out clearly.

It can also be seen in Fig. 3 that the distinction between neutral and double focus is most likely to be signalled by the height of the first pitch peak. As in experiment 1, a small peak on s1 is most likely to signal neutral focus and a large peak is most likely to signal double focus. The boundary between the excursion sizes of the peak on s1 signalling neutral and those signalling double focus is located at about 1.5 E or 6 semitones. There is no corresponding horizontal borderline, which means that the value of p2 may be the same for both neutral and double focus, as was found to be the case of the p2 adjustments in the first part of experiment 1. This is true for most of the values of p2, provided they are smaller than that of p1. The corresponding excursion sizes of the pitch peak on s2 are between 0.5 and 1.5 E or about 2 to 6 semitones. Larger excursion sizes of the second pitch peak are more likely to signal single focus on s2.

In the present experiment, the listeners made a forced choice in every instance, even when the pitch contours sounded somewhat unnatural or when the pitch contours were ambiguous with respect to focus. However, it can be seen in Fig. 3 that most of the pitch contours uniquely correspond to one of the four focus structures. The areas in which the pitch contours were either unnatural or ambiguous are the ones left blank.

6.4 GENERAL DISCUSSION

In experiment 1 we found that, in contrast to the findings in the earlier prominence experiments, the fixed value of one peak did not systematically influence the adjusted value of the other. The different heights of the adjustable pitch peak turned out to be crucial differences between pitch contours when those are signalling various focus structures. On the basis of the results of experiment 1, we hypothesized these heights to be target values for p1 and p2, which would correspond to prototypical pitch contours like the ones shown in Fig. 2. They are also represented by the filled circles in Fig. 3. It can be seen that the prototypes indeed fall in the respective classes of pitch contours signalling the intended focus structures, i.e. our prototypes seem to represent the intended focus structures rather well. In addition, the assumption of target values for p1 and p2 seems to be valid since the separation between different categories of pitch contours turns out to be quite clear, as can be seen in Fig. 3.

In Fig. 3 it can also be seen that the two focus structures which are assumed to signal explicitly unequal prominence of s1 and s2 are clearly separated from each other, one being in the upper left-hand corner, the other one being in the lower right-hand corner. They can be easily discriminated by the listeners. These two areas are separated by the areas in which the pitch contours may signal equal prominence of s1 and s2, viz. the neutral-focus and double-focus structures. The results of experiments 1 and 2 show that, for a wide range of p2 values, the distinction between neutral and double focus is mainly based on the different values of p1, a large value of p1 signalling double focus. In Fig. 3, however, it can be seen that the value of p2 may serve as an additional cue for signalling the intended focus structure. Indeed, the probability of correct discrimination between neutral and double focus is already high if it is based on the discrimination between low and high values of p1, but this probability increases if p2 is taken into consideration as well.

Comparing the results of the present experiments to the grid models, e.g. the one by Horne (1988), it turns out that it makes good predictions for the outcome of our experiments. Obviously, intonation contours for English and Dutch are quite similar with respect to signalling focus.

When we compare the results of the present experiments and those of the previous prominence experiments, the former may clarify some of the issues raised in the latter. One earlier finding is that if s1 and s2 are made equally prominent, the value of p2 is set to be smaller than that of p1, as indicated by the open circles in Fig. 3. This narrowing of the pitch range towards the end of an utterance is generally attributed to declination. The results of the present experiments show that this narrowing of the pitch range is also reflected by the borderline between the pitch contours signalling single focus on s2 and the pitch contours signalling neutral or double focus. Above this borderline, which can be seen to almost coincide with equal values of p1 and p2, the pitch contours signal explicit *unequal* prominence of s1 and s2. Therefore, one may assume that, in the previous prominence experiments, listeners adjusted the value of p2 making it smaller than that of p1 in order to stay on the safe side, i.e. in order to decrease the probability of obtaining unequal prominence.

Another observation made in the earlier prominence experiments was that when s1 and s2 are equally prominent, the difference between the values of p1 and p2 increases with an increase of the value of p1. This is reflected by the fact that the slope of the regression line through the open circles shown in Fig. 3 is less than one. This may be due to the fact that equal low values of p1 and p2 are less likely to result in unequal prominence than equal high values are.
This is also closely related to another observation made by Pierrehumbert (1979) and Terken (1991), viz. that the value of p2 was made even larger than that of p1 in the case of relatively small values of p1. Indeed, the same would be found when extrapolating the trend shown by the open circles in Fig. 3 towards the left-hand side. As can be seen in the lower left hand section of Fig. 3, values of p2 may be larger than those of p1 while listeners will still perceive neutral focus. This means that in the particular case in which the excursion size of the pitch peak on s1 is rather small and the value of p2 is somewhat larger than that of p1, the probability of perceiving *unequal* prominence is still relatively low.
Finally, the observation that pitch contours are classified into categories is in sharp contrast with the findings from the previous prominence experiments in which a scalar relationship between the values of p1 and p2 was found (cf. the open circles in Fig. 3). However, as mentioned before, this difference may be

explained by the difference between the experimental designs, the focus experiments testing at a more abstract level of perception than the prominence experiments. In other words, if subjects are asked to classify pitch contours on the basis of their function, they seem to interpret the prosodic information in a more categorial way, although they may be aware of prominence differences.

6.5 CONCLUSION

In the present experiments we replicated for Dutch the findings for other languages that the perceived focus structure of an utterance very much depends on the properties of its pitch contour. In addition, we conclude that listeners abstract from fine-graded prominence relationships between two pitch peaks in an utterance in determining its focus structure. In brief, we obtained the following specific findings. Single focus on s1 is marked by medium to large values of p1 while s2 should preferably have no excursion size at all. Single focus on s2 is marked by medium to large values of p2 while the peak on s1 is not absent. The value of p1 may even be just slightly smaller than that of p2. Neutral focus is marked by small to medium values of p1 and small to medium values of p2. For very small values of p1, the value of p2 may even be larger than that of p1 without signalling single focus on s2. Double focus is signalled by medium to large values of p1, while the value of p2 is lower than that of p1. A large value of p1 alone may indeed signal double focus even in the case of a relatively small but non-zero value of p2, but a large value of p2 may serve as an additional cue.

More generally, experiment 1 showed how the peak heights of two pitch peaks were related to each other under explicitly different focus conditions. Experiment 2 tested whether listeners are able to recognize the intended focus structure. In both cases, it turned out that listeners have clear ideas about how peak heights signal different focus structures. In Experiment 1, different heights of the fixed peak did not systematically influence the adjusted height of the other peak. In Experiment 2, the results suggested that appropriate values of the pitch peaks are present in the listeners' (and speakers') minds, serving as a kind of target value which should be reached in the case of prototypical pitch contours. Deviant pitch contours, however, are still quite capable of siugnalling the intended focus structure.

APPENDIX

Table A. *Adjusted values of p2 (Hz) under different focus conditions, averaged across the three values of p1 (165, 183, or 202 Hz).*

	intended focus conditions			
	s2	neutral	double	s1
S				
1	206	169	183	110
2	214	209	168	110
3	202	147	185	110
4	206	165	160	115
5	184	152	162	110
6	212	179	167	120
7	197	146	170	119
8	200	126	193	115
9	187	168	157	110
10	214	153	172	110
AV	202	161	172	113

Table B. *Adjusted values of p1 (Hz) under four different focus conditions, averaged across the three values of p2 (143, 160, and 179 Hz).*

	intended focus conditions			
	s2	neutral	double	s1
S				
1	134	169	198	256
2	131	166	267	267
3	152	206	218	251
4	167	179	238	243
5	155	161	174	224
6	155	161	183	242
7	140	182	202	263
8	133	176	191	237
9	151	207	220	249
10	150	141	231	263
AV	147	175	212	249

7 SUMMARY AND CONCLUSIONS

The main conclusion that can be drawn from the experiments described in chapters 2, 3, and 4 is that listeners are very well able to compare the prominence of pitch-accented syllables. The stimuli contained either the same or different kinds of pitch movements: rises, falls, or rise-falls. In comparing prominence, listeners seem to have available two different strategies. One strategy is that they compare pitch-level differences, as is assumed by the Pitch-Level Difference (PLD) model, presented in chapter 3. The other is that they compare the heights of high pitch levels, as is assumed by the High-level (H) model, presented in chapter 4. Which strategy they will use depends on whether the stimuli are in the same or in different registers. The strategy of comparing high pitch levels only works when the stimuli are in the same register.

For stimuli in different registers, Hermes and Van Gestel (1991) had shown that stimuli containing identical pitch movements are perceived as being equally prominent when excursion sizes (or pitch-level differences) are equal in number of ERBs, E. In the experiments described in chapter 3, their results were replicated for stimuli containing different kinds of pitch movements. However, it turned out that excursion sizes of rises in the high register were smaller than expected. It was hypothesized that listeners disliked large excursion sizes for the rise in the high register. A pilot experiment in which rises and falls were compared in different registers, but now with a lower end frequency for the high-register stimuli (and with a smaller step size in the excursion size continuum), gave less clear results than before. Therefore, the stimuli used by Hermes and Van Gestel (1991) and those used in the previous experiments were reanalyzed. The reanalysis revealed that the occurrance of equal excursion sizes measured in E was more or less accurately coupled with an octave relationship between the high pitch levels in the low and high registers, which might have influenced the results. This was checked by a series of experiments which are described in chapter 2. Here it was shown that the conclusion by Hermes and Van Gestel (1991), that equal prominence is perceived in the case of equal excursion sizes in E, still holds when the ratio of the pitch maxima in the low and high registers is smaller or larger than 1:2. It is therefore very unlikely that the outcome of their experiments was influenced by the accidentally occurring octave relationship between their stimuli.

In the second part of chapter 2 we describe another set of control experiments as to the frequency scale of speech intonation. In the previous experiments, the stimuli in the high register were resynthesized from a male voice which was then transposed. The resulting stimuli sounded as if spoken by a boy's or a male falsetto voice. It is possible that the range of such voices is naturally quite narrow. If the listener compensates for this, the use of an original female voice in the high register would lead to equal prominence when the excursion sizes, expressed in E, are larger for the female than for the male falsetto voice. However, the results of the experiments show that the conclusion of Hermes and van Gestel still holds when the stimuli in a pair are resynthesized with different voices, a male and a female voice.

In chapter 3 we have shown that for stimuli in different registers and resynthesized with different kinds of pitch movements, falls have smaller excursion sizes than rises and rise-falls when lending equal prominence. This effect is the same or perhaps even larger when the stimuli are resynthesized in the same low or high register. In the latter case, standard deviations of the measurement points were so small that it made sense to determine the differences between the excursion sizes of rises and falls quantitatively. The outcome of these measurements led to the development of the PLD model, based on the assumption that the perceived prominence, lent by an accent-lending pitch movement, is proportional to the difference measured in E between the pitch level in the accented and that in the previous syllable. The model predicts that two factors will influence the size of the difference S (E) between the prominence lent by rises and that lent by falls. These factors are the rate of pitch declination d (E/s), and the time interval T (s) between the relevant pitch levels. The relevant pitch levels are assumed to coincide with the vowel nuclei of the accented and the previous syllables. The formula for the difference between the excursion sizes of rises and falls that lend equal prominence is $S=2dT$. Increasing the rate of declination d results in an increase of the difference S, as does an increase of T. The experimental evaluation, presented in chapters 3 and 4, shows that the model makes fairly good predictions.

In the second part of chapter 4, however, it was shown that the model's predictions are dependent on the experimental conditions. It turned out that the end frequencies of the baselines in the stimuli play an important role. (One may expect that the declination rates of the baselines are also important, but this experimental condition was not included in our tests). If the end points are not identical but are slightly different from each other, the PLD model no longer makes the right predictions. It turns out that, in that case, perceived equal prominence is no longer based on equality of PLDs in each of the stimuli, but that equality of high pitch levels within the stimulus pair also plays an important, probably even more important, role.

Because of this finding we concluded that the PLD model no longer made the right predictions. Therefore, we proposed an alternative model, the H model, which would account for the measured differences between the excursion sizes of rises and falls lending equal prominence. According to this H model, listeners use the strategy of adjusting equal positions of high pitch levels if the stimuli are in the same register. This alternative model is able to account for the outcomes of the experiments presented in chapters 3 and 4, and the question arises of why this is the case. Surprisingly, it seems to be the case that in those experiments the relevant high pitch levels were located in *un*accented syllables, i.e. before the fall and after the rise (only in the case of the rise-fall is the high pitch level, necessarily, in the accented syllable). So, are we measuring an artefact or are listeners indeed judging prominence on the basis of only high pitch levels? Not quite. In the case of stimuli in different registers, listeners cannot base their judgments on comparison of high pitch levels. Therefore, they must base their prominence judgments on pitch-level differences (chapter 2: equal pitch-level differences in E result in equal prominence). This means that they use some low-pitched reference against which they judge the position of the high pitch level in each of the stimuli. If, in the case of stimuli in the same register, listeners indeed base their prominence judgments on the equality of the positions of high pitch levels, there should also be some low-pitched reference available against which the positions of the high pitch levels are judged. These observations together suggest that when the baselines of stimuli in the same register are almost the same or identical, there is only one low-pitched reference for a pair of stimuli. It is not clear, however, which low pitch levels are chosen to serve as a reference. It could be either the low pitch level before the rise or the low pitch level after the fall. It is sometimes suggested that the end frequency of the baselines are quite constant for a given speaker (Maeda, 1976, referred to in 't Hart *et al.*, 1990, p. 128), so that it may be assumed that the low pitch level in the final syllable of the stimulus containing the fall is the most plausible candidate.

The fact that differences between the high pitch levels seem to be more relevant than those between low pitch levels is in line with the results from other intonation research. For example, Sluijter and Terken (1991), studying paragraph intonation, found that listeners perceived variations in the position of the baseline much less consistently than variations in the position of the topline.

In summary, the results of the experiments presented in chapter 4 point in the direction that prominence perception is indeed based on pitch-level differences, as hypothesized in the PLD model, but that listeners, when judging prominence, make a rough estimate of the positions of relevant pitch levels rather than extracting them from the observable pitch contours. The low-pitched reference level in particular seems to be created on the basis of global pitch information presented in a stimulus pair. In the case of stimuli containing different kinds of pitch movements under the different-registers condition, the results are quite comparable with the results under the same-register condition (chapter 3). But under the different-registers condition it was not possible to draw firm quantitative conclusions because of the relatively large standard deviations.

The conclusion so far is that the PLD model has to be rejected because of its assumption that listeners 'measure' degrees of prominence by exactly determining the pitch interval between the accented and the previous syllable. In addition, we have to conclude that the H model is too simple in assuming that only the equality of high pitch levels is crucial for the perceived equality of prominence lent by different kinds of pitch movements, because this would not explain the findings under the different-registers conditions.

In order to better estimate the contribution to the perception of prominence of pitch *minima* in the stimuli we performed a very small pilot experiment in which two experienced subjects were asked to adjust prominence by manipulating the position of the baseline. It turned out that neither of the subjects performed consistently even at the lowest level of significance, so we decided not to perform any further experiments in that direction. The outcome of the pilot supports the conclusion that for the perception of prominence the exact position of high pitch levels is much more relevant than that of low pitch levels.

A limitation of the present models is that neither the PLD model nor the H model predicts the results obtained in experiments in which two accented syllables are within the same utterance. Apparently, there is a global declination effect working in addition to the local, short-term declination effect which affects pitch-level differences. Due to the effect of long-term declination, two identical accent-lending pitch movements are judged to lend equal prominence when the maximum of the second pitch peak is somewhat lower than that of the first, although their excursion sizes may differ somewhat (Pierrehumbert, 1979; Terken, 1991). In the experiments in which the PLD model was evaluated, this effect of long-term declination was cancelled out by introducing a short pause and a reset of declination lines between the two pitch-accented syllables.

In chapter 5 the effect of (long-term) declination on the difference between pitch peaks on the first and the second pitch-accented syllables (s1, s2) was studied systematically. We studied the relationship between their maxima (p1/p2) and the relationship between their excursion sizes (d1/d2) under different declination conditions of the baseline. In addition, we compared two methods of prominence testing: a single-trial two-alternative forced-choice paradigm in which listeners had to indicate which of the pitch-accented syllables was the most prominent one, and a prominence-adjustment paradigm in which listeners adjusted the excursion size of the second pitch peak making equal the prominence of the first and second pitch-accented syllables. A third comparison was made between the results of stimuli resynthesized in the female pitch range and stimuli in the male pitch range.

We found that the stimuli in the different pitch ranges and the two methods gave essentially the same results. Furthermore, we replicated the findings by Pierrehumbert (1979) and Terken (1991, 1994) that when lending equal prominence, the second pitch-peak maximum should be smaller than that of the first. In addition, it was found that for a given excursion size of the first pitch peak, its prominence increases if the rate of baseline declination increases (p1 higher while d1 constant: p2 higher and d2 larger). Surprisingly, it was also found that for a given value of the first pitch-peak maximum, the results might be interpreted as if its prominence is almost constant even if the rate of baseline declination increases (p1 constant, and therefore d1 smaller: p2 smaller, but d2 constant), when the excursion size d2 is regarded as a rough indicator for the prominence of s2. The results indicate that there is no simple relationship between the values of the pitch-peak maxima, their excursion sizes, and the rate of baseline declination. Moreover, this outcome supports the conclusions from chapter 4 and the hypothesis by Ladd (1993) and Terken (1993), that listeners may use some kind of abstract low- or high-pitched references when judging the heights of pitch maxima and minima in rating prominence. Results of experiments performed by Gussenhoven and Rietveld (1993) indicate that pitch levels at the start of an utterance, if longer than at least two syllables in duration, may be used for establishing such a reference pitch level for the first pitch peak in an utterance (see also Gussenhoven, Repp, Rietveld, Rump, and Terken, 1996). Gussenhoven *et al.* conclude that listeners use a low-pitched abstract reference which is anchored at the start of an utterance and declining at a rate which is independent of the declination rate of the observable baseline.

The conclusions by Gussenhoven *et al.* (1996) give rise to the need for an alternative interpretation of the results presented in chapters 2 to 5. Their most important conclusion with respect to the re-interpretation is that listeners need information about the low pitch levels in at least two syllables before the first pitch-accented syllable when judging the relative height of the (first) pitch peak in an utterance. Since the stimuli in chapters 2 through 4 had only one unaccented syllable before the accented one, this would imply that this is the reason why the changes in the relative height of the baseline did not result in the shifts which were predicted by the PLD model. This might imply that, after all, the PLD model may be a correct model in the case of stimuli with longer onsets before the first pitch-accented syllable. This should be tested with the appropriate stimuli. On the other hand, our experiments have shown that if the information about low pitch levels is less than what is needed for creating a low-pitched reference level, listeners predominantly attend to the pitch peak maximum.

The results presented in chapter 5 support the above re-interpretation, that local pitch information determines the perceived degree of prominence. This strongly implies that listeners have two different strategies for determining the prominence lent by pitch movements available. Firstly, they can use low pitch levels before the pitch peak to determine the relative height of high pitch levels, as in the case of adjusting the prominence of s2. This was also the basic assumption in the PLD model. Secondly, if there is not enough baseline information before the pitch-accented syllable, listeners attend to the high pitch level/pitch-peak maximum in order to determine the prominence of the accented syllable. This may explain the finding in chapter 5 that p1 is directly related to d2. So, a longer onset, including a larger number of unaccented syllables preceding the first pitch peak, might have resulted in a different outcome. The resulting model for prominence perception may then predict a direct relationship between d1 and d2. This should also be tested in future experiments. If a direct relationship between d1 and d2 is indeed found, the need for an abstract declining baseline should then also be questioned again.

The general conclusion to be drawn from the findings reported in chapters 2 to 5 is that listeners are capable of attending to relatively small prominence differences related to pitch phenomena. Excursion-size differences as small as 0.1 E (about 1 semitone) were found to result in statistical significant differences in prominence. This may have at least partly been due to the highly controlled experimental settings. Earlier research (e.g. 't Hart, 1981) has led to the conclusion that differences between excursion sizes larger than three semitones would play a role in the perception of *functional* prominence differences. In chapter 6, we have therefore addressed the functional relevance of such small prominence differences, i.e. their role in the communicative situation. One of the functional aspects of prominence is focusing the listeners' attention on important parts of an utterance. Those parts that get special attention are in narrow focus, those without special attention are in neutral focus. A third condition occurs when they are explicitly out of focus or defocused. In the first part of chapter 6 listeners adjusted the excursion size of one pitch peak, either the first or the second, under the various focus conditions mentioned above. The excursion size of the other peak was kept fixed at three different values.

The results showed that the focus conditions had a highly significant influence on the adjusted excursion sizes. Furthermore, it turned out that the excursion sizes of the fixed peak had no systematic influence on the adjustable excursion sizes. It seems as if the pitch peaks adjusted under the different focus conditions represent some kind of target values. Therefore, our hypothesis is that listeners have quite clear expectations about the structure of focus-related pitch contours. This hypothesis was tested in the second part of chapter 6. In a judgment experiment, pitch contours with different excursion sizes of the first and second peaks were presented and the listeners judged to which of the focus structures the various pitch contours corresponded. It turned out that a given pitch contour corresponded to only one focus structure and that the listeners agreed on that. Large pitch peaks corresponded to narrow focus (either single or double), and small pitch peaks corresponded to neutral focus. When defocused, syllables either had a very small pitch peak, or no pitch peak at all, depending on their position relative to that of the narrow focus in the utterance. The main conclusion from the experiments described in chapter 6 is that the prominence of focus-related pitch peaks is judged globally and in a rather categorial way, which is reflected by the clustering of pitch contours into focus categories. We therefore speculate that, analogous to the vowel categories in an F1/F2-formant space, some category-like areas for combinations of pitch-peak maxima exist for the various focus structures of an utterance. In each of the categories, prototypes represented by the target levels of the pitch maxima are found, which optimally represent the various focus structures. The boundaries between categories may depend on the context in which the utterance is spoken, such as a given F1/F2 combination may be perceived as an /a/ or an /o/, dependent on the context.

This finding of categories containing different pitch contours seems to be in contradiction to the outcome in the earlier chapters, i.e. that prominence differences between pitch-accented syllables are perceived in a gradual way. However, the two sorts of results may be reconciled as follows: listeners decide globally which meaning the utterance conveys, by judging to which of the various focus categories the pitch contour belongs. Additionally, they evaluate the relatively small prominence differences as containing paralinguistic information, for example about the speaker's emotional state. In other words, the two kinds of prominence information, one global and one detailed, seem to contain complementary information which the listener can use in different ways when interpreting the perceived utterance. This dual interpretation of prominence seems to be a very important outcome of the experiments presented in this thesis. Listeners are quite sensitive to small differences in prominence which are related to pitch. The way in which the experiments in chapters 2 to 5 were performed enabled us to measure these relatively small differences. But when listeners have to judge the meaning of utterances, they seem to disregard these relatively small prominence differences which may, however, be relevant for the perception of some paralinguistic aspects of an utterance.

The main conclusion is that the prominence of pitch-accented syllables is closely related to their associated pitch-peak maxima and in particular to the relative height of these maxima in the speaker's register or pitch range. When determining a speaker's pitch range, listeners use information about low pitch levels, commonly occurring in the unaccented syllables. From the results presented in this thesis it has become clear that in addition to this, they may use other information. This might be present in the speaker's voice-source and vocal-tract characteristics.

REFERENCES

Bartels, C., and Kingston, J. (1994). Salient pitch cues in the perception of contrastive focus. In *Focus & Natural language processing, Proceedings of a conference in celebration of the 10th anniversary of the Journal of Semantics, Meinhard-Schwebda, Germany*, Vol. 1, Intonation and Syntax, pp. 1-10.

Beckman, M. E. (**1986**). *Stress and non-stress accent* (Dordrecht: Foris).

Beckman, M. E., and Pierrehumbert, J. (**1986**). "Intonational structure in Japanese and English," in *Phonology yearbook*, edited by Colin J. Ewen and John M. Anderson (Cambridge: Cambridge University Press), Vol. 3, pp. 255-309.

Bolinger, D. L. (**1958**). "A theory of pitch accent in English," Word **14**, 109-149.

Bolinger, D. L. (**1961**). *Generality, gradience, and the all-or-none* (Mouton, The Hague).

Bruce, G. (**1977**). *Swedish word accents in sentence perspective* (Lund: Lund University Press).

Bruce, G. (**1982**). "Textual aspects of prosody in Swedish," Phonetica **39**, 274-287.

Bruce, G. (**1983**). "Accentuation and timing in Swedish," Folia Linguistica **17**, 221-238.

Bruce, G., and Touati, P. (**1992**). "On the analysis of prosody in spontaneous speech with exemplification from Swedish and French," Speech Communication **11**, 453-458.

Cohen, A., Collier, R., and 't Hart, J. (**1982**). "Declination: Construct or intrinsic feature of speech pitch?" Phonetica **39**, 254-273.

Collier, R. (**1970**). "The optimum position of prominence lending pitch rises," IPO Annual Progress Report **5**, 82-85.

Collier, R. (**1991**). "Multi-language intonation synthesis," Journal of Phonetics **19**, 61-73.

Eady, S. J., and Cooper, W. E. (**1986**). "Speech intonation and focus location in matched statements and questions," Journal of the Acoustical Society of America **80**, 402-415.

Eady, S. J., Cooper, W. E., Klouda, G. V., Mueller, P. R., and Lotts, D. W. (**1986**). "Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments," Language and Speech **29**, 233-251.

Fry, D. B. (**1955**). ''Duration and intensity as physical correlates of linguistic stress,'' Journal of the Acoustical Society of America **27**, 765-768.

Fry, D. B. (**1958**). ''Experiments in the perception of stress,'' Language and Speech **1**, 126-152.

Gårding, E. (**1981**). Contrastive prosody. A model and its applications. Studia Linguistica, **35**, 146-165.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hearing Research **47**, 103-138.

Graddol, D. and Swann, J. (**1983**). ''Understanding and describing long term pitch of voice: some physical and social correlates,'' Lang. Speech **26**, 351-366.

Gussenhoven, C., Repp, B. H., Rietveld, A. C. M., Rump, H. H., and Terken, J. M. B. (**1996**). ''The perceptual prominence of fundamental frequency peaks,'' IPO internal report, 1153 (under review for publication in JASA).

Gussenhoven, C., and Rietveld, A. C. M. (**1988**). ''Fundamental frequency declination in Dutch: testing three hypotheses,'' Journal of Phonetics **16**, 355-369.

Gussenhoven, C., and Rietveld, A. C. M. (**1989**). ''Reply to Terken,'' Journal of Phonetics **17**, 365-367.

Gussenhoven, C., and Rietveld, A. C. M. (**1993**). ''Scaling prominence,'' Proceedings Dept. of Language and Speech, University of Nijmegen, **16/17**, pp. 86-90.

Hamon, C., Moulines, E., and Charpentier, F. (**1989**). ''A diphone synthesis system based on time-domain prosodic modifications of speech,'' Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP-89), pp. 238-241.

't Hart, J. (**1981**). ''Differential sensitivity to pitch distance, particularly in speech,'' Journal of the Acoustical Society of America **69**, 811-821.

't Hart, J., and Collier, R. (**1975**). ''Integrating different levels of intonation analysis,' Journal of Phonetics **3**, 235-255.

't Hart, J., Collier, R., and Cohen, A. (**1990**). *A perceptual study of intonation. An experimental-phonetic approach to speech melody* (Cambridge: Cambridge University Press).

Henton, C. G. (**1989**). ''Fact and fiction in the description of female and male pitch,'' Language and Communication **9**, 299-311.

Hermes, D. J. (**1988**). ''Measurement of pitch by subharmonic summation,'' Journal of the Acoustical Society of America **83**, 257-264.

Hermes, D. J., and Van Gestel, J. C. (**1991**). ''The frequency scale of speech intonation,'' Journal of the Acoustical Society of America **90**, 97-102.

Hermes, D. J., and Rump, H. H. (**1994**). ''Perception of prominence in speech intonation induced by rising and falling pitch movements,'' Journal of the Acoustical Society of America **96**, 83-92.

Hill, D. R., and Reid, N. A. (**1977**). ''An experiment on the perception of intonational features,'' International Journal of Man-Machine Studies **9**, 337-347.

Horne, M. A. (**1988**). ''Towards a quantified, focus-based model for synthesizing English sentence intonation,'' Lingua **75**, 25-54.

House, D. (**1990**). *Tonal perception of speech* (Lund: Lund University Press).

Isačenko, A. V., and Schädlich, H.-J. (**1964**). *Untersuchungen über die Deutsche Satzintonation* (Berlin: Akademie-Verlag).

Isačenko, A. V., and Schädlich, H.-J. (**1966**). ''Untersuchungen über die Deutsche Satzintonation.'' In: *Studia Grammatica VII, Untersuchungen über Akzent und Intonation im Deutschen* (Berlin: Akademie-Verlag), p. 7-67.

Kutik, E. J., Cooper, W. E., and Boyce, S. (**1983**). ''Declination of fundamental frequency in speaker's production of parenthetical and main clauses,'' Journal of the Acoustical Society of America **73**, 1731-1738.

Ladd, D. R. (**1993**). ''On the theoretical status of 'the baseline' in modelling intonation,'' Language and Speech **36**, 435-451.

Ladd, D. R., and Terken, J. M. B. (**1995**). ''Modelling intra- and inter-speaker pitch range variation.'' In Proceedings of the XIIIth International Congress of Phonetic Sciences (ICPhS-95), Vol. 2, pp. 386-389.

Ladd, D. R., Verhoeven, J., and Jacobs, K. (**1994**). ''Influence of adjacent pitch accents on each other's perceived prominence: Two contradictory effects,'' Journal of Phonetics **22**, 87-99.

Lehiste, I., and Fox, R. A. (**1992**). ''Perception of prominence by Estonian and English listeners,'' Language and Speech **35**, 419-434.

Leroy, L. (**1984**). *The Psychological Reality of Fundamental Frequency Declination* (Antwerp Papers in Linguistics No. 40, University of Antwerp, Belgium).

Liberman, M. Y., and Pierrehumbert, J. B. (**1984**). ''Intonational invariance under changes in pitch range and length,'' in *Language sound and structure,* edited by M. Aronoff and R. T. Oehrle (Cambridge (MA): The MIT Press), pp. 157-233.

Maeda, S. (**1976**). ''A characterization of American English intonation.'' Doctoral dissertation Massachusetts Institute of Technology, Cambridge, MA.

O'Shaughnessy, D. (**1979**). ''Linguistic features in fundamental frequency patterns,'' Journal of Phonetics **7**, 119-145.

Patterson, R. D. (**1976**). ''Auditory filter shapes derived with noise stimuli,'' Journal of the Acoustical Society of America **59**, 640-654.

Pierrehumbert, J. (**1979**). ''The perception of fundamental frequency declination,'' Journal of the Acoustical Society of America **66**, 363-369.

Pierrehumbert, J. B. (**1980**). ''The phonology and phonetics of English intonation.'' Doctoral dissertation Massachusetts Institute of Technology, Cambridge, MA.

Pierrehumbert, J. B. (**1981**). ''Synthesizing intonation,'' Journal of the Acoustical Society of America **70**, 985-995.

Pierrehumbert, J. B., and Hirschberg, J. (**1990**). ''The meaning of intonational contours in the interpretation of discourse,'' in *Intentions in communication,* edited by Cohen, P. R., Morgan, J., and Pollack, M. E. (The MIT Press, Cambridge, MA), pp. 271-311.

Pompino-Marschall, B. (**1990**). *Die Silbenprosodie. Ein elementarer Aspekt der Wahrnehmung von Sprachrhythmus und Sprechtempo* (Tübingen: Max Niemeyer Verlag).

Repp, B. H., Rump, H. H., and Terken, J. M. B. (**1993**). ''Relative perceptual prominence of fundamental frequency peaks in the presence of declination,'' IPO Annual Progress Report **28**, pp. 59-62.

Rietveld, A. C. M., and Gussenhoven, C. (**1985**). ''On the relation between pitch excursion size and prominence,'' Journal of Phonetics **13**, 299-308.

Rossi, M. (**1971**). ''Le seuil de glissando ou seuil de perception des variations tonales pour des sons de la parole,'' Phonetica **23**, 1-33.

Rossi, M. (**1978**). ''La perception des glissandos descendants dans les contours prosodiques (The perception of falling glissandos in prosodic contours),'' Phonetica **35**, 11-40.

Rump, H. H. (**1992**). ''Timing of pitch movements and perceived vowel duration,'' Proceedings of the 1992 International Conference on Spoken Language Processing (ICSLP-92), pp. 1047-1050.

Silverman, K. E. A., and Pierrehumbert, J. B. (**1990**). ''The timing of prenuclear high accents in English,'' in *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, edited by J. Kingston and M. Beckman (Cambridge: Cambridge University Press), pp. 72-106.

Sluijter, A. M. C. (**1995**). *Phonetic correlates of stress and accent* (The Hague: Holland Academic Graphics).

Sluijter, A. M. C., and Terken, J. M. B. (**1993**). ''Beyond sentence prosody: Paragraph intonation in Dutch,'' Phonetica **50**, 180-188.

Terken, J. M. B. (**1989**). ''Reaction to C. Gussenhoven and A. C. M. Rietveld: 'Fundamental frequency declination in Dutch: testing three hypotheses','' Journal of Phonetics **17**, 357-364.

Terken, J. M. B. (**1991**). ''Fundamental frequency and perceived prominence of accented syllables,'' Journal of the Acoustical Society of America **89**, 1768-1776.

Terken, J. M. B. (**1993**a). ''Synthesizing natural sounding intonation for Dutch: Rules and perceptual evaluation,'' Computer Speech and Language **7**, 27-48.

Terken, J. M. B. (**1993**b). ''Baselines revisited: Reply to Ladd,'' Language and Speech **36**, 453-459.

Terken, J. M. B. (**1994**). ''Fundamental frequency and perceived prominence of accented syllables II: Non-final accents,'' Journal of the Acoustical Society of America **95**, 3662-3665.

Terken, J. M. B., and Van den Hombergh, K. (**1992**). ''Judgments for adjacent and non-adjacent accents,'' in: Proceedings of the 1992 International Conference on Spoken Language Processing (ICSLP-92), pp. 735-738.

Traunmüller, H., and Eriksson, A. (**1995**). ''The perceptual evaluation of *F*0 excursions in speech as evidenced in liveliness estimations,'' Journal of the Acoustical Society of America **97**, 1905-1915.

Välimaa-Blum, R. (**1993**). ''Intonation: A distinctive parameter in grammatical constructions,'' Phonetica **50**, 124-137.

SAMENVATTING

Computerspraak raakt steeds meer ingeburgerd. Toch klinkt de huidige computerspraak niet altijd even goed. Een van de problemen is het toekennen van de juiste mate van prominentie aan geaccentueerde en niet-geaccentueerde lettergrepen. Prominentie van lettergrepen is hun mate van opvallendheid ten opzichte van andere lettergrepen. Onjuiste prominentie beïnvloedt de natuurlijkheid van een gesynthetiseerde uiting, en heeft in het ongunstigste geval tot gevolg dat de betekenis ervan verandert. Een geaccentueerde lettergreep wordt niet alleen gekenmerkt door een betere articulatie, een langere duur en een toename van de luidheid, uit eerder onderzoek is gebleken dat voornamelijk de aan- of afwezigheid van een toonhoogtebeweging bepaalt of hij als geaccentueerd wordt waargenomen of niet. In het voorliggende onderzoek is gekeken naar de invloed van een toonhoogtebeweging op de prominentie van geaccentueerde lettergrepen. Het doel was om een quantitatief model op te stellen voor de bijdrage van toonhoogtebewegingen aan prominentie. Het onderzoek is uitgevoerd in het kader van het intonatie-onderzoek naar de generatie en perceptie van (synthetische) spraak aan het Instituut voor Perceptie-Onderzoek (IPO) in Eindhoven.

Toonhoogte is, zoals gezegd, een eigenschap van spraak die een rol speelt bij accentuering. De waargenomen toonhoogte (Eng.: pitch) wordt bepaald door de frequentie waarmee de stembanden bewegen. Een vrij plotselinge toe- of afname van de frequentie wordt waargenomen als een zogenaamde toonhoogtebeweging. Een spreker maakt hiervan min of meer bewust gebruik wanneer hij een accent wil realiseren.

Accentuering wordt in het dagelijks leven vaak klemtoon genoemd. In de taalkunde wordt met klemtoon (Eng.: stress) echter de fonologische eigenschap aangeduid die aangeeft dat er op een bepaalde lettergreep een accent gerealiseerd kan worden. Bijvoorbeeld, in de zin "Amanda gaat naar Malta" kunnen de lettergrepen /man/ en /mal/ geaccentueerd worden. Welke het uiteindelijk wordt hangt van de context af (en daarmee van de 'focus' in de uiting, zie ook hoofdstuk 6), bijvoorbeeld

a) Wat gaat er gebeuren?
b) Gaat Jan naar Malta?
c) Gaat Amanda naar Cyprus?
d) Gaat Jan naar Cyprus?

Stel dat het zinnetje "Amanda gaat naar Malta" het antwoord is op elk van de vier bovenstaande vragen. In elk van de antwoorden zijn dan de lettergrepen /man/ van Amanda en /mal/ van Malta beklemtoond, maar de focus bepaalt welke lettergreep een accent krijgt. In het geval van voorbeeld a) staat de hele zin in focus en derhalve krijgen beide beklemtoonde lettergrepen een accent. In b) staat Malta expliciet niet, maar Amanda wel in focus zodat alleen /man/ een accent krijgt. In voorbeeld c) is de situatie omgekeerd ten opzichte van b) zodat alleen /mal/ hier een accent krijgt. In d) staan zowel Amanda als Malta in focus zodat zowel /man/ als /mal/ een accent krijgen. Het verschil tussen a) en d) is het verschil tussen een brede, neutrale focus en een dubbele focus.

Welke beklemtoonde lettergrepen in een zin een accent krijgen hangt dus af van de bedoelingen van de spreker. Accent is een fonetische eigenschap, een eigenschap van gesproken taal waarvan je de fysieke grootheden kunt meten. De waargenomen accentsterkte noemen we prominentie. In de hier beschreven experimenten hebben we verschillende eigenschappen van toonhoogtebewegingen onderzocht om hun invloed op prominentie te testen.

Aan het IPO is in het verleden vrij veel fundamenteel intonatie-onderzoek gedaan. Dit heeft onder andere geresulteerd in een model voor het genereren van spraak-intonatie. Een paar belangrijke kenmerken zal ik daar uit lichten om ze even toe te lichten. Een belangrijk kenmerk van neutrale uitingen is dat de toonhoogte aan het begin gemiddeld hoger is dan aan het einde. Dit verloop van de gemiddelde toonhoogte gedurende de uiting heet declinatie. In het Nederlands is het voldoende om twee toonhoogteniveaus aan te geven, een lage en een hoge declinatielijn, hier verder aangeduid als basislijn en toplijn. Om een geaccentueerde lettergreep te realiseren is het in principe voldoende om daarop een vrij abrupte toonhoogverandering tussen de basis- en toplijn te realiseren, zijnde een stijging, daling of een combinatie van een stijging en een daling, een zogenaamde 'punthoed'. De plaats van de beweging en de duur ervan zijn cruciaal voor de accentuering van de juiste lettergreep, de grootte ervan speelt voornamelijk een rol bij het waarnemen van prominentie. Deze rol is het onderwerp van het hier gepresenteerde onderzoek.

Hieronder zal ik kort aangeven hoe de experimenten waren opgezet en wat de belangrijkste uitkomsten ervan waren. De stimuli die gebruikt zijn in de luisterexperimenten beschreven in de hoofdstukken 2 t/m 4 bestonden uit paren van het betekenisloze woordje /ma'mama/, met een accent op de tweede lettergreep. De prominentie van het eerste woord werd steeds gevarieerd door de grootte van de toonhoogtebeweging (excursiegrootte) te variëren en de taak van de proefpersonen was om de prominentie van de geaccentueerde lettergreep van het tweede woord gelijk te maken aan die van het eerste. Zij konden dit doen door de excursiegrootte van de toonhoogtebeweging op het tweede woord in te stellen. In hoofdstuk 2 is gekeken wat de invloed is van het register op de ingestelde excursiegrootte van punthoeden. Met register bedoelen we het gemiddelde frequentiebereik van een stem. Het register van een vrouwenstem ligt ongeveer tweemaal zo hoog als dat van een mannenstem. De vraag is nu wanneer toonhoogtebewegingen evenveel prominentie verlenen aan geaccentueerde lettergrepen die in verschillende registers zijn gesynthetiseerd. Moeten de excursiegroottes gelijk zijn in Hertz, of op een muzikale schaal, bijvoorbeeld in semitonen? Het blijkt dat de punthoeden gelijke prominentie verlenen wanneer de excursiegroottes ervan niet gelijk moeten zijn in Hertz of in semitonen, maar in E. Dit zijn eenheden op de perceptief gedefinieerde frequentieschaal van equivalent-rechthoekige bandbreedtes (ERBs). De ERB-schaal ligt tussen de lineaire Hertz-schaal en de logaritmische semitonenschaal. Het blijkt bovendien niet uit te maken of je voor de synthese van de lage en de hoge stem een originele mannenstem gebruikt, die in het hoge register klinkt als een falsetstem, of dat je voor de lage stem een mannenstem en voor de hoge een vrouwenstem gebruikt.

In hoofdstuk 3 hebben we gekeken of de verschillende toonhoogtebewegingen bij gelijke excursiegrootte evenveel prominentie verlenen. Het blijkt dat stijgingen en punthoeden groter worden ingesteld dan dalingen en dalingen kleiner worden ingesteld dan stijgingen en punthoeden om gelijke prominentie te verlenen. De conclusie is dus dat dalingen meer prominentie verlenen dan stijgingen of punthoeden bij gelijke excursiegroottes. Dit geldt zowel wanneer de beide stimuli in hetzelfde register als wanneer ze in verschillende registers gesynthetiseerd zijn. Deze verschillen bleken niet terug te voeren op een verschil in timing van de beweging ten opzichte van de geaccentueerde lettergreep. Vervolgens hebben we een model opgesteld dat de grootte van de verschillen kan verklaren. Volgens dit Pitch-Level Difference (PLD) model is de waargenomen prominentie recht evenredig met het toonhoogteverschil tussen de geaccentueerde en de voorgaande ongeaccentueerde lettergreep. Volgens dit model is het verschil tussen een stijging en punthoed enerzijds en een daling anderzijds terug te voeren op de mate van declinatie en het tijdsinterval tussen de geaccentueerde en de voorgaande lettergreep (zie ook hoofdstuk 2 Figuur 7). Naar aanleiding van de resultaten van het derde experiment waarin de mate van declinatie systematisch is gevarieerd hebben we geconcludeerd dat het PLD-model vrij goede voorspellingen doet over de samenhang tussen toonhoogteverschillen en prominentie.

In hoofdstuk 4 is een andere factor uit het model getest, het tijdsinterval tussen de geaccentueerde en de voorgaande lettergreep. De resultaten waren echter niet meer vanuit het PLD-model te verklaren. Daarom werd een alternatief model opgesteld dat zowel de nieuwe als de vorige resultaten, voor zover verkregen met stimuli in hetzelfde register, kon verklaren. Volgens dit H model letten luisteraars niet op de grootte van de toonhoogtebewegingen maar stellen zij de excursiegrootte zo in dat het hoge toonhoogteniveau vóór een daling gelijk is aan het hoge niveau van een punthoed of het hoge niveau ná een stijging. Ook dit model hebben we perceptief geëvalueerd door de positie van de lage toonhoogteniveaus in een stimuluspaar ongelijk te maken. Het bleek dat de proefpersonen, wanneer zij gelijke prominentie instelden, inderdaad meer aandacht besteedden aan gelijkheid van de hoge toonhoogteniveaus dan aan de ongelijkheid van excursiegroottes.

De belangrijkste conclusie uit deze proeven is dat hoge toonhoogteniveaus een grotere rol spelen dan lage toonhoogteniveaus bij het bepalen van prominentie. Het zal echter duidelijk zijn dat het H model geen juiste voorspellingen kan doen voor het geval dat de stimuli in verschillende registers zijn gesynthetiseerd. In die situatie blijken toonhoogteverschillen (excursiegroottes) wel degelijk een belangrijke rol te spelen bij de waargenomen prominentie.

In hoofdstuk 5 staan experimenten beschreven waarin we hebben onderzocht wat de invloed van declinatie is op de waargenomen prominentie van twee geaccentueerde lettergrepen in één uiting. De gebruikte stimuli bestonden uit de uiting 'Amanda gaat naar Malta', met /man/ en /mal/ als geaccentueerde lettergrepen. De mate van declinatie is daarbij systematisch gevarieerd en de luisteraars werd gevraagd de excursiegrootte van de punthoed op /mal/ zo in te stellen dat de prominentie ervan gelijk werd aan die van /man/. De resultaten van deze experimenten laten zien dat de grootte van de bewegingen systematisch wordt beïnvloed door de mate van declinatie gedurende de uiting. Bovendien duiden zij erop dat er invloed is van het aantal ongeaccentueerde lettergrepen dat voorafgaat aan de eerste geaccentueerde lettergreep. Vervolgonderzoek dat hier nog niet beschreven is heeft tot de conclusie geleid dat er minimaal twee ongeaccentueerde lettergrepen aan de geaccentueerde vooraf moeten gaan om de luisteraar een goede basis te verschaffen voor het waarnemen van excursiegrootte. Als die basis er niet is, lijkt de luisteraar gebruik maken van de toonhoogte in de geaccentueerde lettergreep, eventueel aangevuld met informatie over het toonhoogtebereik van de spreker. Die informatie zou hij kunnen verkrijgen uit de stemkarakteristiek van de spreker.

In hoofdstuk 6 staan experimenten beschreven met betrekking tot het gebruik van prominentie bij het verlenen van betekenis aan een uiting. Zoals in het begin van de samenvatting al is aangegeven, is een van de belangrijke functies van accentuering het aanduiden van de focusstructuur van een uiting. De focus helpt de luisteraar te bepalen welke onderdelen van een uiting het belangrijkst zijn volgens de spreker. In het zinnetje 'Amanda gaat naar Malta' kan dat dus bijvoorbeeld het woordje Malta zijn, zogenaamde nauwe focus, of de hele uiting (neutrale of brede focus). We hebben luisteraars gevraagd de excursie van de toonhoogtebeweging op Malta of op Amanda zodanig in te stellen dat een van vier mogelijk focusstructuren werd gerealiseerd, nauwe focus op Malta of op Amanda, dubbele focus, of brede focus. Het bleek dat luisteraars dat vrij goed konden en dat de ingestelde waarden niet samenhingen met de excursiegrootte van de toonhoogtebeweging op de andere lettergreep. In het tweede experiment hebben we de uiting met verschillende toonhoogtecontouren aan de luisteraars gepresenteerd en gevraagd welke focusstructuur zij hoorden. Ook deze taak konden zij moeiteloos volbrengen.

Het blijkt dat de resultaten van de beide experimenten goed met elkaar overeenkomen. Bij een brede focus waarbij er geen van beide woorden Amanda of Malta expliciet in focus staat willen luisteraars een relatief kleine toonhoogtebeweging horen, waarbij de absolute piek van de tweede beweging lager moet liggen dan die van de eerste. Een uitzondering hierop is de situatie waarbij de eerste toonhoogtebeweging relatief klein is. In dat geval mag de tweede piek iets hoger liggen dan de eerste. Een dubbele focus, waarbij beide woorden tegelijkertijd in nauwe focus staan, wordt waargenomen wanneer er twee flinke toonhoogtepieken gerealiseerd worden. Ook in deze situatie moet de top van de tweede piek lager zijn dan die van de eerste. Een nauwe focus op Amanda betekent een flinke piek op Amanda maar geen enkele toonhoogtebeweging op Malta. Een nauwe focus op Malta betekent dat er op dat woord liefst ook een flinke beweging moet zijn, maar dat er tevens een toonhoogtebeweging op Amanda moet zijn, ondanks het feit dat dat woord expliciet buiten focus is geplaatst. De top van de eerste piek moet echter steeds beduidend lager liggen dan die van de tweede piek.

Samenvattend kunnen we zeggen dat toonhoogtebewegingen sterk bijdragen aan prominentie en dat de sterkte van de waargenomen prominentie samenhangt met het verschil tussen hoge en lage toonhoogteniveaus in de uiting. We hebben sterke aanwijzingen gevonden dat voor de luisteraar relatief kleine verschillen in de hoogte van de basislijn minder relevant zijn dan kleine verschillen op de toplijn. Als er vrij weinig informatie over de basislijn voorafgaand aan de geaccentueerde lettergreep voorhanden is lijkt de absolute hoogte van hoge toonhoogteniveaus de waargenomen prominentie te bepalen. Het lijkt er op dat de luisteraar dan een schatting maakt van het lage toonhoogteniveau dat bij een bepaalde stem zou horen, om aan de hand van die schatting de grootte van het toonhoogteverschil te bepalen. Bij het maken van de schatting zou een luisteraar zich kunnen baseren op zijn kennis van de persoon/stem van de spreker, of hij zou gebruik kunnen maken van informatie uit het stemgeluid zelf om te schatten wat de grootte van een beweging ten opzichte van het register van de spreker is. Verder is aangetoond dat luisteraars een bepaalde verwachting hebben omtrent de prominentie die samenhangt met de betekenis van een uiting zoals die gerepresenteerd wordt door de focusstructuur. Relatief kleine afwijkingen van die verwachte prominentie lijken vrij goed te worden waargenomen. Zij dragen waarschijnlijk bij aan het al of niet natuurlijk bevonden worden van een intonatiecontour.

LIST OF PUBLICATIONS

Full Papers

Gussenhoven, C., Repp, B. H., Rietveld, A. C. M., Rump, H. H., and Terken, J. M. B. (**1996**). ''The perceptual prominence of fundamental frequency peaks.'' Submitted.

Hermes, D. J., and Rump, H. H. (**1994**). ''Perception of prominence in speech intonation induced by rising and falling pitch movements,'' J. Acoust. Soc. Am. **96**, 83-92.

Repp, B. H., Rump, H. H., and Terken, J. M. B. (**1994**). ''Relative perceptual prominence of fundamental frequency peaks in the presence of declination,'' Inst. for Perception Res., Ann. Prog. Rep. **28**, 59-62.

Rump, H. H., and Collier, R. (**1995**). ''Pitch-peak height and focus,'' Inst. for Perception Res., Ann. Prog. Rep. **30**, 59-62.

Rump, H. H., and Collier, R. (**1996**). ''Focus structures and the prominence of pitch-accented syllables,'' Language and Speech **39**, 1-18.

Rump, H. H., and Hermes, D. J. (**1994**). ''Comparison between two models for prominence perception,'' Inst. for Perception Res., Ann. Prog. Rep. **29**, pp. 29-35.

Rump, H. H., and Hermes, D. J. (**1996**). ''Prominence lent by risisng and falling pitch movements: Testing two models,'' J. Acoust. Soc. Am. **100**, 1122-1131.

Rump, H. H., and Hermes, Dik J. (**1996**). ''Prominence of pitch-accented syllables and the ERB-rate scale.'' Submitted.

Abstracts/Conference Papers

Hermes, D. J., and Rump, H. H. (**1992**) ''Prominence caused by different kinds of pitch movements with different positions in the syllable,'' J. Acoust. Soc. Am. **91**, 2340. Paper presented at the 123rd Meeting of the Acoustical Society of America, Salt Lake City, Utah.

Hermes, D. J., and Rump, H. H. (**1993**). ''The role of pitch in lending prominence to syllables,'' Proceedings of the ESCA Workshop on Prosody, edited by David House and Paul Touati (Working Papers **41**, Lund Univ. Press, Lund, Sweden), pp. 28-31.

Hermes, Dik. J., and Rump, H. H. (**1994**). ''Control experiments on the frequency scale of speech intonation,'' J. Acoust. Soc. Am **96**, 3349. Paper presented at the 128th Meeting of the Acoustical Society of America, Austin, Texas.

Repp, B. H., Rump, H. H., and Terken, J. M. B. (**1993**). "Relative perceptual prominence of fundamental frequency peaks in the presence of declination," J. Acoust. Soc. Am. **94**, 1880. Paper presented at the 126th Meeting of the Acoustical Society of America, Denver, Colorado.

Rump, H. H. (**1992**). "Timing of pitch movements and perceived vowel duration," Proceedings of the International Conference on Spoken Language Processing (ICSLP-92), edited by J. J. Ohala, T. M. Neary, B. L. Derwing, M. M. Hodge, and G. E. Wiebe (University of Alberta, Edmonton, Alberta), pp. 1047-1050.

Rump H. H. (**1993**). "The perception of prominence induced by rising and falling pitch movements." Paper presented at the ELSNET Summer School on Prosody, University College London, July 1993.

Rump, H. H. (**1995**). "Influence of focus structures on tonal targets of pitch peaks," in Proc. of the XIIIth International Congress of Phonetic Sciences (ICPhS-95), edited by Kjell Elenius and Peter Branderud (Stockholm, Sweden), Vol. 3, pp. 664-667.

Rump, H. H., and Hermes, Dik. J. (**1994**). "Some control experiments on a model for prominence perception," J. Acoust. Soc. Am. **96**, 3349. Paper presented at the 128th Meeting of the Acoustical Society of America, Austin, Texas.

Internal reports

Adriaens-Porzig, U., and Rump, H. H. (**1990**). "Das Stilisieren von Timing-Konturen." IPO report, 762.

Rump, H. H. (**1991**). "Timing of pitch movements and perceived vowel duration."IPO report, 812.

Rump, H. H. (**1991**). "Development and perceptual evaluation of a timing module for German diphone speech." IPO report, 829.

Rump, W. (**1989**). "Perzeptieve Evaluierung standardisierter und nicht-standardisierter Diphone in synthetisierten Wörtern und Sätzen." IPO report, 706.

Nawoord

In maart 1989 begon ik aan mijn eerste stage op het IPO, dank zij de bemiddeling van mijn toenmalige mentor Wim Peeters. De prettige en deskundige begeleiding aldaar van Ursula Adriaens-Porzig maakte dat ik er graag een tweede stage aan vast knoopte. Vervolgens mocht ik in deze zeer stimulerende werkomgeving ook nog mijn scriptie-onderzoek verrichten, en wel onder de begeleiding van dé intonoloog van Nederland, Hans 't Hart.
Blijkbaar was de gunstige ervaring wederzijds want na het afronden van mijn studie lag er een AIO-projektvoorstel waarvoor ik als uitvoerder werd aangezocht, een verzoek waaraan ik gaarne voldeed en waarvan nu het resultaat voorligt.
Graag wil ik mijn promotor René Collier bedanken voor zijn grote bereidheid mee te denken en te discussiëren over de richting van mijn onderzoek. Vooral het feit dat hij ondanks zijn vele drukke andere werkzaamheden bijna elke week wel een uurtje voor me vrijmaakte gaf me een zeer goed gevoel. Mijn co-promotor Dik Hermes wil ik van harte bedanken voor zijn practische hulp, en zijn bereidheid om me elk moment van de dag met goede raad terzijde te staan. Zijn vele stimulerende en uitdagende opmerkingen hebben voor een zeer groot deel het verloop van mijn onderzoek bepaald.
Verder wil ik graag iedereen bedanken die het me mogelijk heeft gemaakt om mijn onderzoek af te ronden. Mijn dank gaat daarbij vooral naar mijn IPO-collega's die bijna allemaal weleens als proefpersonen aan de soms moeilijke en vervelende, maar gelukkig nooit erg langdurige, luisterexperimenten belangeloos hebben willen meedoen. Daarnaast was de technische en administratieve ondersteuning perfect. Voor hun hulp bij statistische kwesties dank ik Huib de Ridder en Jan Ruijter.

Verder dank ik mijn ouders en schoonouders, verdere familie, mijn vrienden, en mijn nieuwe collega's voor hun belangstelling voor mijn promotie-onderzoek.

Tenslotte wil ik mijn partner Hannie bedanken, zonder wie ik nooit aan dit werk had kunnen beginnen noch het tot een goed einde had kunnen brengen.

Curriculum Vitae

De auteur van dit proefschrift werd geboren op 17 januari 1960 te Bremen, Duitsland. Hij groeide op in Groningen alwaar hij achtereenvolgens De goudvink, de J. H. Topschool en het Willem Lodewijk Gymnasium doorliep. In 1987 begon hij in Utrecht met de studie Algemene Letteren. Na het met lof behalen van de propaedeuse, begon hij im 1988 met de kopstudie Fonetiek. Na het doctoraal examen startte hij in 1991 op het Instituut voor Perceptie-Onderzoek in Eindhoven met het onderzoek waar dit proefschrift een verslag van is.

In 1995 is hij begonnen met de deeltijdopleiding tot bibliothecaris aan de Hogeschool van Amsterdam (IDM: informatiedienstverlening en -management). Sinds juni 1996 is hij werkzaam bij de bibliotheek van de Universiteit Utrecht.

STELLINGEN

behorende bij het proefschrift
H. H. Rump: *Prominence of pitch-accented syllables*

I

Het toonhoogteverschil waarop luisteraars hun oordeel over de prominentie van een lettergreep baseren is niet gelijk aan de excursiegrootte van een toonhoogtebeweging. Dit toonhoogteverschil kan, afhankelijk van de stimuli, kleiner, gelijk, of zelfs groter zijn dan de excursiegrootte.

II

Gelijke prominentie van geaccentueerde lettergrepen van mannen en vrouwen wordt waargenomen wanneer, onder verder identieke omstandigheden, de excursiegroottes van de toonhoogtebewegingen gelijk zijn op de perceptieve frequentieschaal van equivalent-rechthoekige bandbreedtes (ERBs).

III

De waargenomen grootte van een toonhoogtebeweging wordt bepaald door de toonhoogte in de geaccentueerde lettergreep ten opzichte van de toonhoogte in de voorgaande ongeaccentueerde lettergrepen mits dat er minimaal twee zijn.

IV

Een luisteraar realiseert zich pas dat prominentie bijdraagt aan de herkenning van de betekenis van een zin wanneer er met de prominentie iets mis gaat. Het is als de bibliothecaris die pas wordt gemist wanneer hij er niet is.

V

Het beluisteren van steeds hetzelfde woord heeft zoals bekend een vervreemdend effect waardoor de betekenis er van wordt losgekoppeld. Dit effect kan worden omgedraaid door uit te gaan van een woord zonder betekenis. Na enige tijd zal een proefpersoon er een eigen betekenis aan gaan hechten.

VI

Privatisering van het reizigersvervoer zal op den duur leiden tot slechtere service in gebieden buiten de randstad, waardoor de verstedelijking in die gebieden zal toenemen.

VII

Om energie te besparen worden de ramen van nieuwbouwwoningen in Nederland steeds kleiner waardoor zij meer en meer op die van Zuideuropese huizen gaan lijken. Deze vorm van Europese eenwording is funest voor de in Nederland nog bloeiende kamerplantencultuur.

VIII

Het zou de kwaliteit van het fietspadennet in Nederland ten goede komen wanneer degenen die verantwoordelijk zijn voor aanleg en onderhoud ervan zelf eens op de fiets zouden stappen.

IX

Huizinga signaleerde in 1919 reeds de tendens van een afnemende leeslust omdat 'wij ons steeds meer van het lezen naar het kijken gewend hebben' (Herfsttij der Middeleeuwen, p. 254). Deze voorliefde voor kant en klare beelden is zo sterk dat zelfs het intellectuele niveau van de beelden op de televisie heden ten dage de mensen er nog niet toe kan brengen eens naar een goed boek te grijpen.

X

Het aantal publikaties tussen 1965 en 1987 is ongeveer even groot als het aantal publikaties vóór 1965 (Informatie en Informatiebeleid 8, 1990, p. 68). Het gemiddelde aantal lezers van een wetenschappelijk artikel is heden hooguit 5 (Noel Malcolm, Prospect, zomer 1996). Gezien de verwachting dat de stijging van het aantal publikaties zich zal doorzetten schiet publiceren op den duur zijn doel voorbij want blijft als enige lezer de auteur.