Timing of pitch movements and accentuation of syllables in Dutch

Citation for published version (APA):

DOI:
10.1121/1.419623

Document status and date:
Published: 01/01/1997

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Timing of pitch movements and accentuation of syllables in Dutch

Dik J. Hermes
Institute for Perception Research/IPO, P.O. Box 513, NL 5600 MB Eindhoven, The Netherlands

(Received 1 February 1996; revised 6 February 1997; accepted 26 May 1997)

In this study, the relation between the timing of a rising or falling pitch movement and the syllable it accentuates is investigated. The five-syllable utterance /mamamama/ was provided with a relatively fast rising or falling pitch movement. The timing of the movement was systematically varied and Dutch subjects were asked to indicate which syllable they perceived as accented. In order to find out where in the pitch movement the cue which induces the percept of accentuation is located, the duration of the pitch movement was varied. In order to find out which segments of the utterance this characteristic is linked to, the duration of the /m/ was varied. The results showed that the percept of accentuation is induced by a change in pitch at the start of the movement. The moment at which the course of pitch starts to change significantly determines which syllable is perceived as accented. If this moment lies some tens of milliseconds before the P-center, i.e., the perceptual moment of occurrence of the syllable, the preceding syllable is perceived as accented. For a rise, a high accent is perceived; for a fall, a low accent. If the pitch change occurs after this moment, the syllable with this P-center is perceived as accented. For the rise, a low accent is then perceived; for the fall, a high accent. This will be discussed in the light of earlier research on accentuation and of theoretical knowledge about pitch accents. © 1997 Acoustical Society of America. [S0001-4966(97)04309-9]

PACS numbers: 43.71.An, 43.71.Es, 43.70.Fq [WS]

INTRODUCTION

This paper is concerned with the relation between an accent-lending pitch movement and the syllable accentuated by this pitch movement. Two topics are addressed. The first topic has to do with the question of which syllable is perceived as accented as the temporal position of the pitch movement in an utterance is varied. To answer this, subjects were presented with a polysyllabic nonsense word with a pitch contour containing a rise or a fall, and they were asked to indicate the syllable they perceived as accented. The second topic deals with what kind of accent is perceived by the listener. To answer this, subjects were asked to indicate whether they perceived a high or a low accent, as the temporal position of a pitch movement was varied.

The first topic deals with the same question as addressed by Van Katwijk and Govaert (1967), who varied the position of a rising or a falling pitch movement within a synthetic four-syllable speech sound /hikxyxyl/. They asked subjects to rate the “level of accentuation” of the various syllables. They found that for the second, third, and fourth syllables higher levels of accentuation were attained when they were accented by a rise than when accented by a fall. The first syllable reached a higher level of accentuation when accented by a fall than when accented by a rise, but this could be attributed to a very strong declination. In the stimulus synthesized by Van Katwijk and Govaert (1967), pitch frequency decreased from 160 to 110 Hz in 1200 ms, the duration of the stimulus. This amounts to 42 Hz/s or 0.45 oct/s, about twice as much as is normal in neutral speech. Because of this strong declination the pitch in the utterance with the fall started at a very high level which probably caused the high ratings for the level of accentuation of the first syllable. The second conclusion by Van Katwijk and Govaert was that the “pitch falls exert their prominence-lending force later than do pitch rises” (p. 115). This was based on the observation that falls in positions where they induced the highest level of accentuation of a syllable were positioned later in the syllable than rises in positions where they induced the highest level of accentuation. These experiments were carried out at a time when good speech synthesis was not yet possible. Furthermore, it is hard to decide from these experiments what attribute of the pitch contour induces the accentuation of a syllable and to what speech segment this attribute is linked.

Another related experiment is reported by Hasegawa and Hata (1992). They varied not only the position of a falling pitch movement in the nonsense utterance /mamama/, but also the steepness of the fall, which was completely situated within the vowel of the second syllable. They asked Japanese subjects to indicate whether they perceived an accent on the first or on the second syllable. They concluded, firstly, that the second syllable was only perceived as accented when the accenting fall started a certain distance into the vowel and, secondly, that, the steeper the fall, the more the subjects tended to indicate the first syllable as accented.

Much more is known about the temporal position of a pitch movement within a syllable and the kind of accent perceived. In various descriptions of intonation, pitch movements are divided into different categories on the basis of their position in the syllable. For example, in the description of Dutch intonation as given by ’t Hart et al. (1990), two kinds of accent-lending rises are distinguished: an early rise starting before the vowel onset and a late rise starting after the vowel onset. Hill and Reid (1977) obtained a similar...
result for English. They presented subjects with a pair of utterances which were identical except for the timing of a rising pitch movement. For a specific timing difference between the two pitch movements, subjects classified the utterances as more different when the two pitch movements started on different sides of the vowel onset than when they started both on the same side of the vowel onset. In the description by 't Hart et al. (1990), a realization of an early rise is indicated with '1,' a late rise with '3.' In autosegmental phonology, accents lent by these two types of pitch movements correspond with L+H* and L*+H, respectively (Pierrehumbert, 1980).

For the accent-lending fall there is less agreement as to the division into different categories. 't Hart et al. (1990) give only one phonetic category of full-sized, accent-lending falls: the ‘A.’ Gussenhoven and Rietveld (1992/1993) distinguish two phonological categories, for Dutch, too, which they indicate with H*L and L*H. The second category is distinguished from the first by the presence of down-step (Pierrehumbert, 1980), a phenomenon for which there is no exact equivalence in the description by 't Hart et al. (1990).

In the autosegmental phonology of English intonation (Pierrehumbert, 1980; Beckman and Pierrehumbert, 1986) two bi-tonal falls are distinguished, H*+L and H+L*. In Swedish, two lexical tones are distinguished. Bruce (1977) showed that timing differences between the falling pitch movements involved played a crucial role in distinguishing between the two.

Thus, various investigations have been carried out into the relation between the timing of a pitch movement and the kind of tonal event that is perceived. In the investigations reported in this paper, this problem also comes into play in the third experiment reported here, but in the first two experiments the main question is: which syllable is perceived as accented as the temporal position of a pitch movement is varied? Experiments are reported in which the timing of rising and falling pitch movements was systematically varied and Dutch subjects were asked to indicate which syllable they perceived as accented. The location in the utterance where the percept of accentuation shifts from one syllable to the next will be referred to as the accentuation boundary. In order to find out where the cue for accentuation of the pitch movement is located, at its onset or at its offset, pitch movements with three different durations were used. Furthermore, the duration of the speech segments was varied in order to investigate to which segment the attribute of the pitch contour which induces the accentuation of the syllable is coupled. The rationale of these two setups is given in the description of the experiments.

I. EXPERIMENT I

In this experiment, pitch movements of different durations were used: 80, 120, and 160 ms. The onset of these pitch movements was systematically varied over the utterance /mamamamama/. Both rises and falls were used. When the timing of the pitch movement is gradually shifted to later in the utterance, the percept of accentuation will, at a certain location in the utterance, shift from one syllable to the next. The moment at which this occurs is the accentuation boundary.

A. Rationale

This experiment was carried out with pitch movements of three different durations in order to find out whether the cue for accentuation is at the onset or at the offset of the pitch movement. The rationale underlying this experimental setup is illustrated in Fig. 1. A hypothetical accentuation boundary between two syllables /ma/ is indicated by the dotted vertical lines in the lower panels of Fig. 1(a), (b), (d) and (e), and in the oscillograms in Fig. 1(c) and (f). The basic idea is that the location of the cue for accentuation at the accentuation boundary will be more or less independent of the duration of the pitch movement. So, if the pitch cue which induces the percept of accentuation is located at or
immediately after the onset of a pitch movement, the location of the onset of the pitch movement at the accentuation boundary will be more or less independent of the duration of the pitch movement. For the rise, this situation is depicted in Fig. 1(a), and in Fig. 1(b) for the fall. The location of the pitch cue is indicated by the ellipses. The position of this pitch cue is not exactly at the turning point of the pitch movement, since the perceptual process which determines where the cue is located will have an integration time, and this integration time may be somewhat longer for stimuli with weaker cues. It is supposed that the onset of the longer-duration pitch movements is less abrupt and, therefore, weaker than that of the shorter-duration pitch movements. The onset of the pitch movements at the accentuation boundary may, therefore, shift somewhat to the left. The offset of the pitch movement at the accentuation boundary will accordingly shift to the right as the duration of the pitch movement gets longer. This is shown in the lower panels of Fig. 1, where the position of the onsets of the pitch movements of three different durations at the accentuation boundary are indicated by circles and the position of the offsets of these pitch movements are indicated by crosses. If, on the other hand, the cue which accentuates the syllable is at or immediately after the offset of the pitch movement, the offset of the pitch movement at the accentuation boundary will be more or less independent of the duration of the pitch movement. This situation is illustrated in Fig. 1(d) for the rise, and in Fig. 1(e) for the fall. The hypothetical location of the pitch cue at the end of the pitch movement is again indicated by the ellipses. As argued before, the perceptual process which determines the moment at which the pitch movement has been completed will have an integration time. Therefore, the offset of the pitch movement at the accentuation boundary will at most shift a bit to the left as the duration of the pitch movement gets longer, but certainly not to the right. The onset of the pitch movement at the accentuation boundary will accordingly shift much more to the left. This is shown in the lower panels of Fig. 1(d) and (e), where the position of the onsets of pitch movements of three durations at the accentuation boundaries are indicated by circles and the position of the offsets of these pitch movements are indicated by crosses.

B. Method

1. Stimuli

The stimuli were derived from a natural three-syllable utterance /mamama/, spoken with an accent on the second syllable. Speech modifications were carried out with pitch-synchronous-overlap-add (PSOLA) techniques (Hamon et al., 1989) with which pitch and time modifications can be applied to speech signals without affecting the speech quality too much. The first step was to triplicate the middle syllable, resulting in the utterance /mamamamama/ consisting of five syllables, of which the middle three were identical as to amplitude and spectral envelope. Next, the original pitch contour was replaced by a rising or a falling pitch movement superimposed on a declination line. The rises and falls had durations of 80, 120, or 160 ms. A stimulus continuum was created by shifting the pitch movement through the utterance in steps of 20 ms. The onset of the first element of the continuum was at the start of the stimulus. The offset of the last element was at the end of the stimulus. The declination rate was fixed at 0.7 E/s, where E is nr of ERB (Hermes and Van Gestel, 1991; Glasberg and Moore, 1990). In the male-voice range used in this study this amounts to about 24 Hz/s or 0.39 oct/s. The interval between successive onsets of the pitch movements in the continuum was 20 ms. The utterance lasted 1.32 s, so there were 62 stimuli with rises of 80 ms and 62 with falls of 80 ms, 60 stimuli with rises of 120 ms and 60 with falls of 120 ms, and 58 stimuli with rises of 160 ms and 58 with falls of 160 ms. Stimuli with rises and falls were mixed. There were 180 rises and 180 falls, so there were 360 different stimuli.

The order of presentation was random. The set of 360 stimuli was presented twice to each subject in the same random order. These stimuli were presented in six sessions of 120 stimuli. In each session, each set of 120 stimuli was preceded by 12 practice stimuli, during which the subjects were allowed to adjust the intensity of the stimuli to a comfortable loudness level. This set of practice stimuli approximately covered the range of the stimuli presented in the actual experiment.

2. Subjects and procedure

Eleven subjects took part in this experiment; all were students or research associates of the Institute of Perception Research. Some of them had experience in speech perception experiments, but there was no clear relation between the presence of this experience and the results. Their age ranged from about 20–40 years. All were native speakers of Dutch. Each subject did the test individually in an interactive computer session, seated in a quiet room. The subject started the presentation of a stimulus by pressing the return key on the keyboard or by clicking a button on the computer screen. He/she could listen to a stimulus as often as desired, by pressing or clicking repeatedly. The task was to indicate on which of the five syllables an accent was perceived. Thus, there were five response classes corresponding to whether an accent was perceived on the first, the second, the third, the fourth, or the fifth syllable. All responses, including the number of times a subject listened to each stimulus, were recorded.

C. Results and discussion

The results for the stimuli with the rises are presented in Fig. 2. The range of the onsets of the rises is shown in Fig. 2(a) for the stimuli with the 120-ms rise. For all five response classes, distributions were calculated of the onsets of the corresponding pitch movements. These distributions are presented in Fig. 2(b). They represent, for each response class, the number of these responses as a function of the onset time of the pitch movement. The upper part of the panel shows the results for the stimuli with the pitch movement of 80 ms, the middle part for the pitch movements of 120 ms, and in the lower part of the pitch movements of 180 ms. The onset distribution of the rises accenting the first
syllable is indicated with a thicker line than the other distributions. It can clearly be seen that, when the pitch movement starts early in the utterance, in almost all cases the first syllable is perceived as accented. However, after about 200 ms, more and more subjects start to indicate the second syllable as accented. Then, when the onset of the rise is later than about 250 ms, the listeners indicate in almost all cases the second syllable as accented. At about 450 ms there is a shift to the third syllable, etc. So, clear intervals can be seen where the corresponding syllable is almost always selected as the accented syllable. These regions are indicated by the numbers 1 to 5. What is most important at the moment are the locations in the speech signal where the percept of accentuation changes from one syllable to the next. These are the category boundaries on the continuum of onset and offset times before which a pitch movement accents one syllable and after which it accents the next. These are the moments in the speech signal where the distributions cross. They were calculated after smoothing the distributions, since without such smoothing the statistical fluctuations on the distributions could strongly influence the result. The results are presented in Fig. 2.

FIG. 2. Determination of the rises at the accentuation boundaries. In (a) the stimulus continuum is presented for the 120-ms rises. In (b), for each of the five response classes, the frequency distributions are presented of the onset times of the pitch movements accenting the first syllable. The numbers in the plots indicate the syllable which, according to these distributions, is most often indicated as accented. The thick continuous line is the distribution of the onset times of the pitch movements accenting the first syllable. The lowest panel shows the oscillogram of the /mamamamama/ stimulus monotonized at 100 Hz; the dotted lines are the transitions from the /a/’s to the /m/’s; the dashed lines are the transitions from the /m/’s to the /a/’s; the diamonds are the average of the onsets of the three rises at the accentuation boundary.

FIG. 3. Determination of the falls at the accentuation boundaries. Results are presented in the same way as in Fig. 4.

The results for the stimuli with the falls are presented in Fig. 3 in the same way as for the rises. In Fig. 3(a) the stimulus range of the falls is presented for the 120-ms fall. In Fig. 3(b) the onset distributions of the falls accenting each of the syllables are indicated with a thicker line than the other distributions. It can clearly be seen that, when the pitch movement starts early in the utterance, in almost all cases the first syllable is perceived as accented. However, after about 200 ms, more and more subjects start to indicate the second syllable as accented. Then, when the onset of the rise is later than about 250 ms, the listeners indicate in almost all cases the second syllable as accented. At about 450 ms there is a shift to the third syllable, etc. So, clear intervals can be seen where the corresponding syllable is almost always selected as the accented syllable. These regions are indicated by the numbers 1 to 5. What is most important at the moment are the locations in the speech signal where the percept of accentuation changes from one syllable to the next. These are the category boundaries on the continuum of onset and offset.
the five syllables are presented. As for the rises, clear intervals can be distinguished where in almost all cases either the first, the second, the third, the fourth, or the fifth syllable is perceived as accented. The locations where these distributions cross supply the accentuation boundaries of the syllable. The falls at these accentuation boundaries are presented in Fig. 3(c), in the lower part of which the onsets and offsets are presented. As for the rises, the onsets of the falls at the accentuation boundaries shift somewhat to the left, but the offsets shift to the right as the pitch movement gets longer. This is the situation presented in Fig. 1(b), which shows that, as for the rises, the cue for accentuation of the fall is located at the onset of the fall.

In Figs. 2(b) and 3(b), the onset distribution of the pitch movements accenting the first syllable is indicated by a thicker line than the onset distributions of the other syllables. Especially in the case of the fall, it can be seen that pitch movements starting very late in the utterance often induce the percept of an accent on the first syllable. This is probably due to the weakness of the accenting cue of such a late pitch movement. As has been shown elsewhere (Hermes, 1995), subjects tend to indicate the first syllable as accented in the absence of a clear accenting cue.

II. EXPERIMENT II

The previous experiment showed that the temporal location of the change in pitch at the onset of the pitch movement determined which syllable was perceived as accented. Inspection of Fig. 2(c) and (d) and Fig. 3(c) and (d) showed that this change in pitch was located close to the syllable onset (cf. Caspers and Van Heuven, 1993; Van Santen, 1995), about 90 ms before the vowel onset. To find out whether this segmental anchor point was indeed the vowel–consonant (VC) transition or some other point in the utterance, the duration of the /m/ was varied. In one set of stimuli the duration of the /m/ was halved, in the other it was doubled.

A. Rationale

The rationale of this experiment is presented in Fig. 4. The left three panels present the situation if the accentuation boundary is coupled with the syllable onset, i.e., the VC transition. This VC transition is indicated with the vertical dotted line, while the hypothetical accentuation boundary is represented by the continuous vertical line. In Fig. 4(a) three rises of different duration are depicted at the hypothetical accentuation boundary for the stimulus with the short /m/. In Fig. 4(b), these three rises are presented for the normal /m/, while in Fig. 4(c) these three rises at the hypothetical accentuation boundary are shown for the stimulus with the long /m/. Since in these three left panels the situation is depicted in which the cue for accentuation is coupled with the syllable onset, the distance between the syllable onset and the accentuation boundary, indicated by the double arrow in Fig. 4(a)–(c), remains constant, while the distance between the accentuation boundary and the vowel onset varies.

In the right three panels of Fig. 4, the situation is sketched in which the hypothetical accentuation boundary is coupled with the vowel onset, i.e., the consonant–vowel (CV) transition. The hypothetical accentuation boundary is again indicated by the continuous vertical line, while the CV transition is indicated with the dashed vertical line. In Fig. 4(d)–(f), the three accent-lending rises at the accentuation boundary are presented for the stimuli with the short /m/, the normal /m/, and the long /m/, respectively. Since in these three right panels the situation is depicted in which the cue for accentuation is coupled with the vowel onset, the distance between the accentuation boundary and the vowel onset, indicated by the three double arrows in Fig. 4(d)–(f) remains constant, while the distance between the VC transition and the accentuation boundary varies.

So, the aim of this experiment was to find out whether the accentuation boundary is coupled with the syllable onset or the vowel onset. To this end, accentuation boundaries
were determined for /mamamamama/ stimuli with /m/’s of three different durations. If the accentuation boundary is linked up with the syllable onset, the distance between syllable onset and accentuation boundary will remain constant. If, on the other hand, the accentuation boundary is linked up with the vowel onset, the distance between vowel onset and accentuation boundary will remain constant.

B. Method

1. Stimuli

In order to restrict the set of stimuli, only the accentuation boundary between the second and the third syllable was investigated in this experiment. Hence, only stimuli were used in which the second or the third syllable was perceived as accented. For the stimuli with short /m/’s, the range of the onsets of the pitch movements is depicted in Fig. 5(a) for the rises and in Fig. 5(e) for the falls. For the stimuli with the short /m/, the onsets ranged from 260–460 ms after the onset of the utterance with intervals of 20 ms. The pitch movements again had a duration of 80, 120, and 160 ms, and there were rises and falls, resulting in a total of 66 stimuli with short /m/’s. In the stimuli with the long /m/’s the onsets of the pitch movements ranged from 460 to 780 ms after the onset of the utterance, again with intervals of 20 ms. The pitch movements had durations of 80, 120, or 160 ms, and there were rises and falls, resulting in a total of 102 stimuli with long /m/’s. The oscillograms of the stimuli are presented in Fig. 5(d) and (h) for the short-/m/ stimuli, and in Fig. 6(d) and (h) for the long-/m/ stimuli. The durations of the three /m/’s in the original /mamama/ were about 70, 80, and 70 ms. The duration of these /m/’s were doubled and halved by means of PSOLA techniques, as were the triplicating of the second syllable and the pitch manipulations.

No subject complained about any unnaturalness of the stimuli, except that two subjects with a long experience of PSOLA-manipulated speech signals could identify the speech signals as manipulated with PSOLA. One of them heard a weak irregularity somewhere in the stimulus, probably due to some failure of PSOLA. It is assumed that this did not have a relevant influence on the results. Since excluding their data did not change the results, their data are included in the following description.

2. Subjects and procedure

The same 11 subjects were used as in the first experiment. Their task was to indicate whether they perceived an accent on the second or on the third syllable of the utterance. Stimuli with rises and falls were mixed as in Experiment I, but the stimulus set with the short /m/’s was presented in a separate session from the stimulus set with the long /m/’s. Both sets of stimuli were again presented twice in the same random order. Each time, each set was preceded by 12 practice stimuli covering the range of the stimuli presented in the actual experiment.

C. Results

The results are presented in Fig. 5 for the stimuli with the short /m/ and in Fig. 6 for the long /m/. Since the stimulus range was limited to the stimuli accenting the second or the third syllable, only the accentuation boundaries between the second and the third syllable are presented. The results are similar to those obtained in Experiment I, except that the accentuation boundaries are positioned at different moments. For the stimuli with the short /m/’s, the accentuation boundary was located at about 65 ms before the vowel onset. For the stimuli with the long /m/’s, this was at 110 ms before the vowel onset. It is worth noting, as can be seen in Fig. 6, that this position is within the /m/. The results for the three sets of stimuli are summarized in Fig. 7. The left three panels present the rises at the accentuation boundaries and the oscillogram of the stimuli, for the

FIG. 5. Accentuation boundaries between the second and the third syllable for the /mamamamama/ stimuli in which the duration of the /m/ was halved. In panels (a)–(d) the results for the rises are plotted, in panels (e)–(h) the results are for the falls. Panels (a) and (e) give the stimulus range of the rise and the fall of 120 ms. In (b) and (f) the onset distributions of the pitch movements of the three different durations are presented. The thicker line indicates the frequency distribution of the number of responses in which the subject indicated the second syllable as accented; the thinner line gives the number of times the third syllable was perceived as accented. The points where these two distributions cross yield the onsets of the pitch movement at the accentuation boundaries. The pitch movements at these accentuation boundaries as determined in this way are presented in panels (c) and (g), in which the lower part gives the onsets (small circles) and the offsets (crosses) of the pitch movements at the accentuation boundary. The two lowest panels present the oscillogram of the /mamamamama/ stimulus monotonized at 100 Hz; the diamonds are the average of the onsets of the three pitch movements at the accentuation boundary.
short /m/ at the upper panel, for the normal /m/ at the middle panel, and for the long /m/ at the lower panel. The falls at the accentuation boundaries are similarly presented in the right three panels. If these results are compared with the two situations sketched in Fig. 4, it can be concluded that the accentuation boundary is certainly not linked up with the syllable onset. For the short-/m/ and the normal-/m/ stimuli, the accentuation boundary is close to the syllable onset, but for the long-/m/ stimuli, the syllable boundary precedes the accentuation boundary by at least 60 ms. On the other hand, the distance between the accentuation boundary and the vowel onset, though less variable, is not really constant either. For the short /m/, this distance is about 65 ms, for the normal /m/ 90 ms, and for the long /m/ 110 ms. This will be discussed in more detail in the Discussion, where it will be argued that the accentuation boundary is neither strictly coupled with the syllable onset nor the vowel onset, but with the perceptual moment of occurrence of the syllable, the so-called P-center (Morton et al., 1976).

III. EXPERIMENT IIIa

In this experiment, the main question was not which syllable was perceived as accented, but what kind of pitch accent was perceived. As described in the Introduction, in Dutch as well as in British English intonation, two kinds of rises are distinguished, an early rise starting before the vowel onset and a late rise starting after the vowel onset. For the fall, the situation is less clear. Here, this issue was addressed by asking subjects to indicate, for a number of different timings and for three different durations of the pitch movement, whether they perceived a high or a low accent. If a pitch movement occurs early in the interval where it accents a syllable, an accent lent by a rise is perceived as a high accent, while an accent lent by a fall is perceived as a low accent. If the pitch movement occurs late in the interval where it accents a syllable, a rise induces a low accent, while the fall induces a high accent. The location in the syllable where this transition occurs will be referred to as high–low boundary. This was examined both for the rise, where indeed the percept changes from high to low, and for the fall, where the percept actually changes from low to high.

FIG. 6. Accentuation boundaries between the second and the third syllable for the stimuli in which the duration of the /m/ was doubled. Results are presented in the same way as in Fig. 4.

FIG. 7. Summary results of experiment II. Pitch movements of three different durations at the accentuation boundaries are presented together with the corresponding oscillograms of the stimuli monotonized at 100 Hz. In (a)–(c) the results for the rises are presented, while in (e)–(g) the results for the falls are shown. In the upper two panels (a) and (d) the results with the short /m/ are presented, in the middle two panels (b) and (e) are those with the normal /m/, and in the lower two panels (c) and (f) are those with the long /m/. The diamonds are the average of the onsets of the three pitch movements of different duration at the accentuation boundary.
The stimuli consisted of a restricted set of the stimuli used in Experiment I. The location of the onsets of the pitch movements ranged from 480–680 ms after the physical start of the utterance. This range, shown for the pitch movement of 120 ms in Fig. 8(a) and (e), was chosen in such a way that in almost all cases an accent was perceived on the third syllable. For some of the earliest pitch movements some subjects indicated having perceived an accent on the second syllable. Subjects were explicitly asked to concentrate on the third syllable and indicate whether they perceived a high or a low third syllable. The interval between the elements of the continuum was 20 ms, so there were 11 rises, just as there were 11 falls. There were rises and falls of three durations: 80, 120, and 160 ms, resulting in 66 different stimuli.

2. Subjects and procedure

The same 11 subjects took part as in the previous two experiments. Their task now was not to indicate on which syllable they perceived an accent, but to indicate whether they heard a high or a low accent. The series of 66 stimuli was presented twice in the same random order. Each time, this series of 66 stimuli was preceded by 12 practice stimuli approximately covering the range of the stimuli used. During the presentation of these practice stimuli, subjects had the opportunity to adjust the volume of the stimuli to a comfortable loudness level.

B. Results and discussion

The results are shown in Fig. 8. As in the previous task, the boundary between the categories was quite well defined, showing that the subjects were capable of doing the task. The onset distributions of the accent-lending pitch movements are presented in Fig. 8(b) for the rises, and in Fig. 8(f) for the falls. The pitch movements at the high–low boundary as determined by the onset distributions are presented in Fig. 8(c) and (g). Just as for the accentuation boundary, it appears that the onset of the pitch movements at the high–low boundary shifted somewhat to the left, while the offset shifted to the right. This shows that here, too, the location of the onset of the movement determined whether a high or low accent was perceived.

In the previous experiments concerned with the accentuation boundary, these boundaries were at about the same position in the syllable for both the rises and the falls. The high–low boundaries for the rises, however, appeared to be later than those for the falls. The average of the onsets of the three rises is indicated with a diamond in Fig. 8(d); that of the onsets of the three falls in Fig. 8(h). The onsets of the falls at the high–low boundary were located some 20 ms after the vowel onset; those of the rises 45 ms later, some 65 ms after the vowel onset. Based on some comments by subjects, and by informal listening, the possibility was considered that this may be attributed to a systematic bias to identify an accent lent by rise with a high accent and an accent lent by fall with a low accent. Especially in the neighborhood of the high–low boundary the listener will hesitate between indicating a high or a low accent. As a consequence, the effect of this bias on the listener will be strong especially in this neighborhood. Hence, if, in this transition region, accents lent by rises are systematically indicated as high accents, the high–low boundary for the rise will show a bias to the right. If, similarly, in the neighborhood of the high–low boundary accents lent by falls are systematically indicated as low accents, the high–low boundary for the fall will also show a bias to the right. Hence, the difference between the high–low boundary of the rise and the fall cannot be explained by such a bias if this bias is equal for the rise and the fall.

FIG. 8. Determination of the high–low boundary of the third syllable of the /mamamamama/ stimulus. The stimulus range comprised those positions in which the third syllable is perceived as accented. For the 120-ms rise, this range is presented in panel (a), and for the 120-ms fall in panel (e). Distributions of the onset times of the pitch movements lending a high or a low accent to the third syllable are presented in panel (b) and (f). In (c) and (g), the pitch movements are presented at the high–low boundary as determined on the basis of the onset distributions. The two lowest panels present the oscillogram of the /mamamamama/ stimulus monotonized at 100 Hz. The diamonds are the average of the onsets of the three pitch movements of different duration at the high–low boundary.
fall. Since the high–low boundary for the rise comes later than that of the fall, the bias should be larger for the rise than for the fall. It was felt that this bias may have been introduced by mixing the stimuli with the rises and the falls within one experimental session.

IV. EXPERIMENT IIIb

This experiment examined whether mixing stimuli with rises and falls within one experimental session might have introduced a systematic bias on the part of the subjects to classify an accent lent by a rise as a high accent and an accent lent by a fall as a low accent. Therefore, the same stimuli were presented, but this time rises and falls were presented in separate sessions.

A. Method

1. Stimuli

The stimuli presented in this experiment were exactly the same as in the previous experiment. There were 33 rises and these were presented twice in one experimental session. There were 33 falls, also presented twice, now presented in a separate session. So, the stimulus set of the rises and that of the falls both contained 66 stimuli. Within the set of rises the presentation of the stimulus was random, just as in the set of falls. The session with the rises was preceded by 12 practice stimuli approximately covering the range of the stimuli, all consisting of rises. The session with the falls was also preceded by 12 practice stimuli approximately covering the range of this set of stimuli, all consisting of falls.

2. Subjects and procedure

Twelve subjects participated, none of whom participated in the previous experiment. For six subjects, the stimulus set with the rises was presented first, while for the other six subjects the set with falls was presented first. Their task was again to indicate whether they heard a high or a low accent.

B. Results and discussion

The results are presented in Fig. 9, along with the results from the earlier experiment. The onset times of the rises at the high–low boundary are presented in Fig. 9(a), those for the falls in Fig. 9(b). It can be seen that, relative to the mixed experimental condition, the high–low boundary for the rises is about 35 ms later when the session with the rises is presented first, and that it is about 35 ms earlier when the falls are presented first. The size of the effect is about as large for the rises of 80, 120, and 160 ms. For the falls, the accentuation boundary hardly changes with experimental condition. This shows that confusion of rises with high accents is a stronger effect than confusion of falls with low accents. It is, however, difficult to decide whether the effects are significant. If they are, and if indeed the difference in the high–low boundary between rises and falls is a consequence of an asymmetrical bias to classify accents lent by rises as high accents and accents lent by falls as low accents, it is difficult to explain why this effect is larger for the subjects who first carried out the session with the rises. Perhaps the session with the rises made the subjects aware of the relevant perceptual difference in this task, thereby avoiding the confusion of a high accent with an accent lent by a rise in the following experimental sessions. No other explanation could be found.

Observe furthermore that in all three conditions, the high–low boundary still comes later for the rises than for the falls. Whether this residual difference is still due to this confusion of high accents with accents lent by rises or is systematic is discussed later on.

What is not problematic is the conclusion that the location of the onset of the pitch movement determines whether a high or a low accent is perceived, just as it determines which syllable is perceived as accented. The timing of the offset is irrelevant.

V. GENERAL DISCUSSION

A. Large effects

The main conclusion of these experiments is that both the cue that determines which syllable is perceived as accented and the cue indicating whether a low or a high accent is perceived are located at the onset of the pitch movement. In all the results, the position of the onsets of the pitch movements at the category boundaries remain about constant or shift somewhat to the left as the pitch movement becomes longer. The offsets of the pitch movements at the category boundaries shift to the right. This situation was sketched in Fig. 1(a) and (b). This virtually excludes the possibility that the location of the offset of the pitch movement provides the cue for accentuation and the high–low distinction between accents. (In principle, it is still possible that a pattern recognition process in which the offset is taken into account plays a role. However, this would then require a complicated lookahead strategy involving a delay of at least some 80 ms.)

The second important conclusion from these experiments, though less definite and in need of further experimentation, is concerned with the location of the category boundaries within the utterance. Based on the results described in this paper and on a review of the literature, it is argued that the temporal relation between the onset of the pitch movement and the perceptual moments of occurrence of the syllables, the so-called P-centers (Morton et al., 1976), determines whether the pitch movement accents the syllable and what kind of accent is perceived.

The P-center is closely related to the vowel onset of a syllable, though three phenomena show that the P-center and the vowel onset cannot be identified with each other. The first two phenomena have to do with the relation of the
P-center with the segmental structure of the syllable (Marcus, 1981). First, as the consonant cluster preceding the vowel has a longer duration, the P-center comes earlier with respect to the vowel onset. Second, as the syllable rhyme, i.e., that part of the syllable which follows the vowel onset, has a longer duration, the P-center comes later with respect to the vowel onset. The third phenomenon is that the P-center depends on the kind of pitch movement realized on the syllable. Janker and Pomponio-Marschall (1991) found a difference in P-center location of Thai words with a falling lexical tone and words with a rising lexical tone. The first of these phenomena will be linked with the shift to the left with respect to the vowel onset as the /ml’s become longer. The third will be linked with the difference in the high–low boundary of falls and rises as found in experiment III.

The accentuation boundaries as found for the syllables with /ml’s of different durations are located at 65 ms before the vowel onset for the stimuli with an /ml halved in duration, at 90 ms for the stimuli with an /ml of original duration, and at 110 ms for the stimuli with an /ml doubled in duration. These shifts were unexpected, but they are exactly in the range of shifts in the perceptual center (P-center) found when varying the duration of the syllable onset (Marcus, 1981). Pomponio-Marschall (1989) varied the duration of the /ml in the syllable /ma/ and found about the same kind of shift as found in this study.

The category boundary of the high–low distinction for the fall is located 20 ms after the vowel onset. For the rise, this boundary is located 35–60 ms later, depending on the experimental context. If the experimental sessions with the falls preceded those with the rises, the difference between the high–low boundaries for the rises and the falls is smaller than when rises and falls are mixed within a session, and when the sessions with the rises are presented before those with the falls. So far there is no good explanation for this. Based on the results of experiment IIIb, it is suggested that listeners have a tendency to identify accents lent by rises with high accents. This may not, however, explain everything. Moreover, in this respect there is also a clear correspondence with results found for P-centers. This also relates to the finding by Janker and Pomponio-Marschall (1991) of a difference in the P-center of Thai words with a falling lexical tone and words with a rising lexical tone. These rising and falling tones are both late in the syllable, so the correspondence between these results and those described here is remarkable. It is felt that these issues need further experimentation. As a last remark, it is tempting to identify the timing of the onset of the pitch movements found in this study with the timing of the “start of the accent command” in Fujisaki’s two-component theory of intonation (Fujisaki and Hirose, 1984; Fujisaki, 1993).

It is interesting to compare these conclusions with results obtained by House (1990). He found that sensitivity for pitch changes is reduced in speech segments of significant spectral change. Naturally, the vowel onset, or its presumed perceptual correlate, the P-center, lends itself prominently to the timing of syllable events. If it is assumed that, some tens of milliseconds before the P-center, sensitivity to pitch change is reduced, we must conclude that for early rises and falls, where the onset of the movement is located in this interval of reduced sensitivity, a jump to a pitch level different from what is expected on the basis of only declination will be perceived. So it will be this jump which induces the percept of accentuation. Still following House (1990), if the change in the course of pitch occurs after the vowel onset, so after the interval of reduced sensitivity to pitch change, we naturally perceive the change itself which then provides the listener with the cue for accentuation.

In summary, it now appears that two different mechanisms can induce the percept of an accent. If the change in pitch occurs at a location such as the P-center of the syllable where we are not very sensitive, the percept of a jump in pitch in the following vowel induces the accent. If the change in pitch takes place at a location where it is readily perceived, the pitch change itself induces the accent.

B. Small effects

The onsets of the pitch movements at the accentuation boundaries most often show a small shift to the left as the pitch movements become longer. This shift shows that the interval between the onset of the pitch movement and the moment after this onset at which the listener decides that an accent-lending change in the course of pitch has occurred, is longer for the longer pitch movements. This can be explained quite naturally, because, the excursion sizes of the different pitch movements being equal, the slope of the shorter pitch movements is differs more from the slope of declination than from the slope of the longer pitch movements. So, the change in the course of pitch is more abrupt for the shorter pitch movements. Therefore, the integration time of the perceptual process which decides whether an accent-lending change in the course of pitch has taken place is up to 40 ms longer for the longer pitch movements. What cannot be explained is that for the rises this shift is generally less than for the falls. Inspecting the data shows that for the rises this shift varies between −6 and 26 ms, and for the falls between 24 and 42 ms. Why this shift with longer pitch movement duration is in general a bit larger for the falls than for the rises still remains unclear. In general, the change in the slope of the pitch contour is more abrupt for the rise than for the fall, as a consequence of the falling declination. This should be coupled with a later category boundary for all falls, however, and cannot in itself explain why the shift with duration of the pitch movement is larger for the falls than for the rises.

C. Relation with previous studies

As mentioned in the Introduction, only a few previous studies have been carried out on the accentuation boundaries between two syllables. The oldest one is the study by Van Katwijk and Govaert (1967), who also varied the position of a rising or a falling pitch movement. Their first conclusion was that the rise was more effective in accenting a syllable than a fall. This conclusion is not clearly substantiated by the results obtained here, but this can be explained by the different experimental setup used by Van Katwijk and Govaert. First, the declination in their experiments was almost twice
as strong as the declination used in the experiments described here. The consequence of this is that the contrast in slope between declination and accent-lending fall becomes less. In addition, as the authors write, an utterance with such a strong declination starts so high that the subjects may perceive a “virtual rise” at the start. This virtual rise is a natural filling up of the so-called “hat pattern,” a sequence of a rise and a fall, which is very common in Dutch intonation. Furthermore, in a pilot experiment preceding the experiments described here (Hermes, 1995), it also appeared that for the /mamamamama/ stimuli used here, subjects most often indicated the first syllable as accented in the absence of any pitch movement besides declination.

The second conclusion by Van Katwijk and Govaert was that pitch falls exert their prominence lending force later than pitch rises. This cannot be explained by these differences in experimental setup. If, due to the strong falling declination, the fall becomes less conspicuous, the time to perceive the change in the course of pitch should be longer for the fall than for the rise. However, if this is the case, the accentuation boundary of the falls should be earlier than that of the rises. The easiest solution is then to attribute the difference to the bad synthesis quality of the stimuli used by Van Katwijk and Govaert. This may have interrupted the natural temporal relations of the vowel and the noise parts of the stimuli, but it is felt that here, too, more experimentation with syllables with fricative consonants as used by Van Katwijk and Govaert are needed.

Collier (1970) carried out an experiment in which he asked subjects to position an accent-lending rise in an utterance at the point where it lent optimum prominence to a specified syllable. He concluded that “subjects tend to relate its position to the vowel onset and not to the beginning of the syllable” (p. 83), and “The comparison of the location of pitch rises of different lengths is preferably made by considering the position of their tops” (p. 84). This last conclusion was also based on the observation in a pilot experiment that the distribution of the onsets of the pitch movements showed a much greater spread than the distribution of their peaks. This might seem contradictory to our results, but it follows from the different task of Collier’s subjects. He asked subjects to adjust the timing of the pitch movement in such a way that it lent optimal prominence to a specified syllable. This may amount to adjusting the timing of the pitch movement in such a way that the peak of the rise fell in the vowel. For the longer rises, this may have the consequence that their starts were located in the vowel of the previous syllable and, as Collier mentions, “the preceding syllable might become slightly stressed, thus weakening the optimal stress impression on the next, indicated one” (p. 84). Thus, Collier observed that in the adjusted position of the longer pitch movements the accent could shift to the previous syllable.

This, of course, raises the important issue of the relation between prominence, a gradual property of the syllable, and accentuation, a binary property of the syllable. Furthermore, the fact that a pitch movement may not lend an accent at all has been ignored. The experiment was deliberately carried out with /mamamamama/ stimuli in which the middle three syllables were identical in all respects except for the pitch movement realized on them. In the original three-syllable utterance the middle syllable carried an accent, and, consequently, had the temporal and loudness properties of a stressed syllable (Sluijter, 1995). As has been argued (Hermes, 1995), in the absence of a clear accenting cue, the first syllable is mostly indicated as accented. Since this is not observed in the experiments reported here, all pitch movements used apparently lent an accent.

The results reported here may contradict results found by Hasegawa and Hata (1992). They used a three-syllable utterance /mamama/, which was provided with a pitch contour consisting of a gradual rise up to the vowel of the second syllable, an accent-lending fall starting within the vowel and a more gradual declination up to the end of the utterance. They concluded that the second syllable was only perceived as accented when the accenting fall started after the vowel onset. The main difference with our stimuli was that the first syllable had a gradually rising pitch contour, which in itself might have been a cue for an accent on the first syllable. Furthermore, if subjects have to choose between the first and the second syllable, they may have shown a bias for the first.

Some more literature is available on the high–low distinction between pitch accents. The results described in Experiment III corroborate a finding obtained by Bruce (1977) for Swedish word accents. In one of Bruce’s experiments, he varied the timing of a fall in such a way that somewhere between the two extremes of the continuum, the word accent and with it the meaning of the utterance changed. The duration of the fall varied between 40, 60, and 80 ms. He asked subjects to indicate what they heard. Bruce also found that the onset of the falls at the category boundary between the two Swedish word accents shifted to the left as the pitch movement became longer. The durations used by Bruce, 40, 60, and 80 ms, were smaller than the durations used here, 80, 120, and 160 ms, but they all agree in the sense that the perceptual decision was apparently taken between 20–50 ms after the onset of the pitch movement, and that this period was shorter for more rapid pitch movements. Bruce also found that the location of the category boundary was in the neighborhood of the vowel onset. For Swedish word accents, this boundary separates acute accent I, in the terms of this paper a fall lending a low accent, and the grave accent II, in the terms of this paper a fall lending a high accent. This conclusion by Bruce must perhaps be modified somewhat on the basis of results from Engstrand (1995). After analyzing a limited set of production data, Engstrand concludes that “one—and probably the—positive requirement on the word accent contrast is an F0 fall on the primary stressed vowel in grave words. If the acute accent is at all associated with some such positive requirements, it remains to be specified” (p. 178). Translated into the results of this study, he explains that the grave accent II in Swedish is positively characterized as a late fall within the primary stressed vowel, while the acute accent I does not show this characteristic.

As to the difference between rises lending a high accent and rises lending a low accent, the results described in this paper are in agreement with the results found by Hill and Reid (1977) for British subjects. As described in the Introduction, they found that the boundary between the two cat-
categories was approximately at the vowel onset.

Van Dommelen (1995) found that, in Norwegian, the fall, in contrast to the rise, brought about a perceptual lengthening of the vowel of almost 20 ms. In fact, another interpretation would be that in the stimulus with the fall the P-center is about 20 ms earlier. If the perceived start of a vowel occurs at the P-center, the fall will last longer by the same amount as the P-center starts earlier. What is still unexplained is that Van Dommelen did not find the same lengthening of the vowel containing a fall in German syllables presented in sentence context.

More results on the category boundary between high and low accents, as lent by falls for Dutch, are provided by Gussenhoven and Rietveld (1992/1993). They showed that the boundary between these two categories is not exactly the vowel onset, but depends on the segmental structure of the onset and the rhyme of the syllable. They reported not to have found a direct relation with the P-center as determined by the algorithm by Pomponio-Marschall (1989, 1991). These P-centers were not determined by a perception experiment, however, and inspection of their results shows that the locations of the category boundaries shifted in the same way with different syllable onsets as P-centers.

In most of the literature described, only separate rises or separate falls were taken into account. In many languages, an accent can also be lent by a combination of a rise and a fall, the rise–fall. At least in Dutch intonation, the location of these rise–falls is such that the rise starts about 50 ms before the vowel onset, while the fall starts about 80 ms after the vowel onset. The results of this study have been argued to indicate that an accent can be induced by two different mechanisms, a jump to a new pitch level in the syllable nucleus, and a change in pitch within the syllable nucleus. In terms of the results of this paper, this shows that rise–falls contain both cues for accentuation. The position of the rise effects that, in the following nucleus, a jump in pitch is perceived which induces a high accent in that syllable, while the onset of the fall within the syllable nucleus is also a cue for a high accent on this syllable. This may explain why the rise–fall is an effective cue for accentuation. In Hungarian the situation is more complicated. In this language, the timing of the accent-lending rise–fall on the last accented syllable determines whether a question or a statement is perceived (Gósy and Terken, 1994). The interpretation of these results in terms of the results described here is complicated by the presence of two movements. The same complication in this respect applies to results presented for German by Kohler (1991).

D. Conclusions

Combining all the evidence, it now appears that the moment at which the change in the course of pitch at the onset of a pitch movement is perceived determines which syllable is perceived as accented. If this moment occurs some tens of milliseconds before the P-center of the syllable, the syllable preceding the syllable with this P-center is perceived as accented. When it comes later, the syllable with this P-center itself is perceived as accented. If the change in the course of pitch at the onset of the pitch movement occurs before or at the P-center, a high accent is perceived for the rise and a low accent for the fall. If the change in the course of pitch at the onset of the pitch movement occurs after the P-center, a low accent is perceived for the rise and a high accent for the fall.

Combined with the results obtained by House (1990), it is suggested that the perception of a pitch accent can arise from two different mechanisms. In the first, the perception of a change in the course of pitch at the onset of the pitch movement induces the perception of an accent. In the second, the perception of an accent is brought about by a pitch level jump between syllable nuclei. The perception of a pitch change gives rise to the perception of an accent on the syllable with the rhyme in which the pitch change is perceived. For a rise this induces a low accent, while for the fall it is high. If a pitch jump is perceived, the latter of the two syllables between which the jump takes place is perceived as accented. In this case the rise induces a high accent, while the fall induces a low accent.

ACKNOWLEDGMENTS

I would like to thank Jacques Terken, Frédéric Beau-gendre, and Marc Swerts for their critical reading of previous versions of this manuscript. I would also like to thank Jacques Terken for his suggestion for the third experiment to ask subjects to indicate whether they heard a high or a low accent, and David House for correcting my original view on Experiment IIIb.


Dik J. Hermes: Timing ofpitch movements