Journal of Phonetics (1995) **23**, 375–385

Aligning pitch targets in speech synthesis: effects of syllable structure

T. Rietveld and C. Gussenhoven*
*Department of Language and Speech, and * Department of English, University of Nijmegen, Erasmusplein 1, NL-6525 HT Nijmegen*

Received 28th February 1994, and in revised form 21st March 1995

The perceptual cross-over point between a Dutch "flat-hat" contour with an early final fall and one with a late final fall is sensitive to the segmental structure of the syllable accented by the fall. The findings are interpreted in terms of segmental influences on the alignment of accented targets (H*): voiced segments in the syllable onset and coda pull it leftwards and rightwards, respectively, while voiceless segments in the onset also exert a leftward pull. The P-centre as defined in the work of Pompino-Marschall does not explain the variation observed. It is concluded that phonetic implementation rules align the intonational targets with reference to the beginning of the syllable, but additionally allow the voiced segments in the onset and rhyme to centre the location to a certain extent within the voiced stretch. © 1995 Academic Press Limited

1. Introduction

Current expectations of the quality of synthetic speech are not merely concerned with the intelligibility of the synthesis product. Synthesized speech needs to simulate prosodic features in a variety of segmental conditions. It is becoming increasingly clear that the relation between intonational features and the segmental conditions is more complex than has previously been thought. In particular, recent phonetic research has shown that the alignment of pitch targets of intonation contours is sensitive to factors both within the accented syllable as well as in subsequent portions of the utterance (Silverman & Pierrehumbert, 1990). House (1989) has shown that the duration of the f_0-movements in accent-lending peaks in British English correlates with the length of the voiced section of the syllable rhyme. Both Steele (1986) and House (1989) found a tendency for the locations of the beginnings of pitch movements to vary with syllable length and the number of following unaccented syllables.

In this article the question arises whether the systematic variation in alignments that has been reported in production data is reflected in the listener's perceptual strategies. In the computer implementation of the description of Dutch intonation by Gussenhoven (1988, 1991) and Van den Berg, Gussenhoven & Rietveld (1992), described in Gussenhoven & Rietveld (1992), the target of the accented tones H* and L* is insensitive to the segmental composition of the accented syllable: the alignment point is defined as falling a certain percentage of the total vowel duration after the beginning of the vowel. It is encoded in the variable STARTIME, a

0095-4470/95/040375 + 11 $12.00/0

multiplication factor running from 0.01 to 0.99 which is applied to the total vowel duration; the result represents the distance between the vowel onset and the pitch target. This procedure was intended to simulate the finding by Steele (1986) that longer syllables tend to have later alignments of pitch peaks. However, this procedure appears to be too crude. With STARTIME = 0.10, we have found that the identity of the intonation contour appeared to change when we changed the text on which the contour was realized. This conclusion is also reached by van Santen & Hirschberg (1994) on the basis of production data. It is self evident that an adequate speech-synthesis system should reliably produce the same intonation pattern for any given intonation input, irrespective of the segmental conditions.

A preception experiment is reported in which the effect of the segmental composition of the syllable on the perceptual cross-over point between early and late pitch falls was investigated. The intonational contrast used was that between a downstepped and a non-downstepped "flat hat". Before presenting the details of the experiment, that nature of this contrast and the way in which it has been accounted for in different analyses is discussed.

1.1. *An intonational contrast*

As the research conducted in the "Dutch school" of intonation analysis has shown ('t Hart, Collier & Cohen, 1990), the timing of pitch movements is a crucial ingredient in the identity of Dutch intonation contours. Fig. 1 shows artificial versions of two contours, both of which mark two syllables as accented, corresponding to the positions of the rise and the fall. In contour (a) (in the top portion of the figure), a falling pitch movement occurs in the accented syllable. By contrast, the final falling movement in contour (b), is earlier than that in contour (a), starting before the accented syllable instead of during it.

The contrast between the early and late fall also occurs in English and German. Pierrehumbert (1980) analysed the English contour (a) as %L H* H*L L%, and contour (b) as %L H* HL* L%. In this notation, % is a diacritic for a boundary tone, while the asterisk marks the tone that lines up with the accented syllable. In

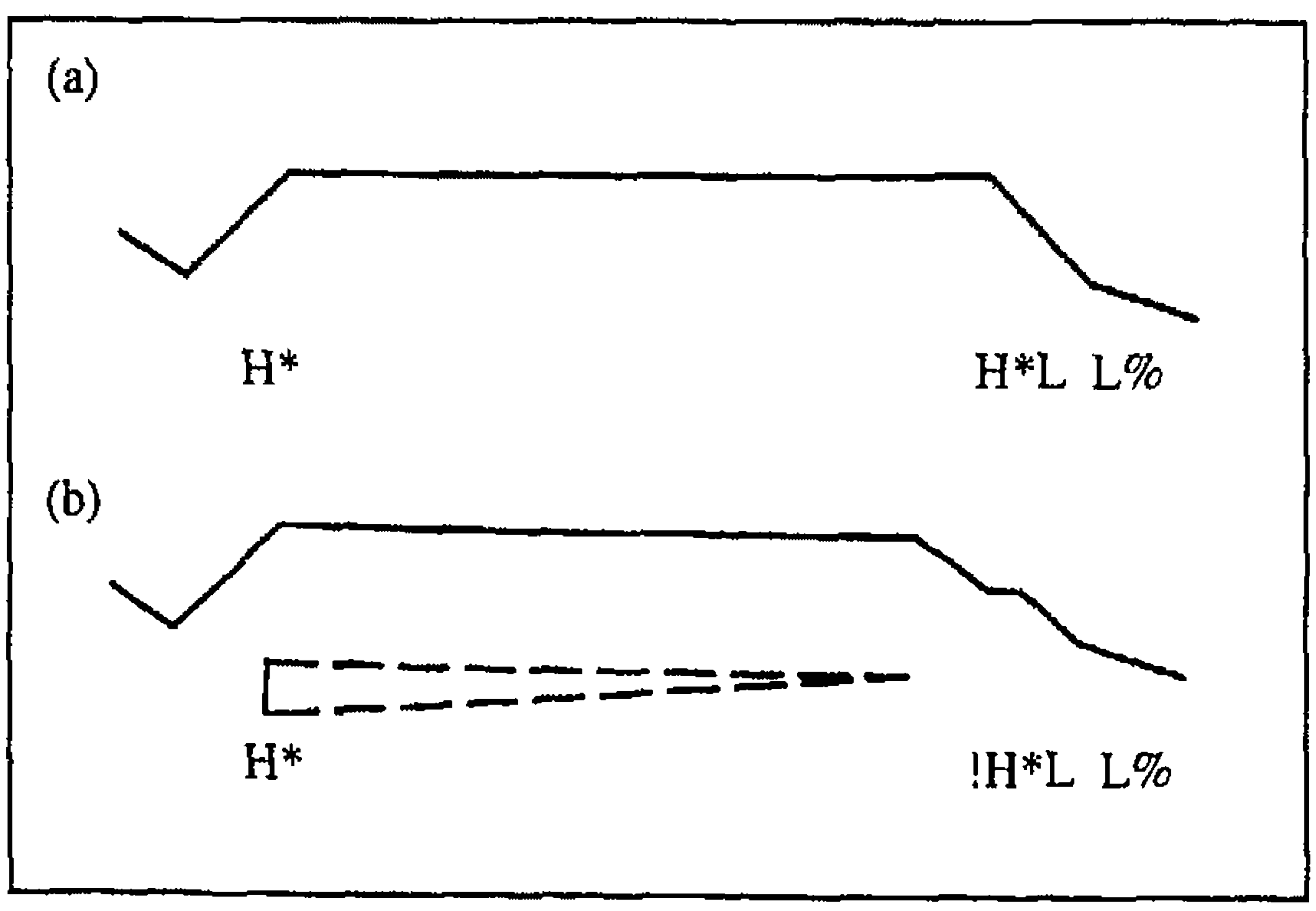

Figure 1. Two artificial contours as produced by the synthesis programe described in Gussenhoven & Rietveld (1992). Contour (a) is a non-downstepped "flat hat" contour, and contour (b) a downstepped one.

her analysis, the L* of the HL* pitch accent is realized as a lowered (downstepped) H (in narrow transcription given as "!H*"), and this L* is therefore equivalent to the !H* in the transcription of contour (b) in Fig. 1. The contrast is akin to that between an early and a central accent-lending peak in German described by Kohler (1990:115ff).

't Hart *et al.* (1990) did not recognize the contrast between early and late falls as categorical, specifying both contours as "1∅A" (where "1" represents an accent-lending rise, ∅ is the high level pitch connecting the peaks, and "A" the accent-lending fall). Their movement "A" was described as occurring late in the accented syllable, but in the audio-taped illustrations (Collier & 't Hart, 1981) as well as other work (e.g., de Zitter, 1992; Caspers & Van Heuven, 1993) the fall began before the accented syllable. In computer implementations of their intonation grammar (Collier, 1991), the early placement was used for the "1∅A", while the late placement was reserved for the "pointed hat" contour (a single pitch peak on a single accented syllable, symbolized in their notation as "1&A" (cf. Collier, 1991: Table II). Thus, the different timings of the "A" movement were interpreted as contextual variants of the same movement in the traditional analysis of Dutch intonation represented by 't Hart *et al.* (1990). More recently, however, the need to include two distinct categories of falling movements of the type illustrated in the contours in Fig. 1 in the grammar of Dutch intonation has been recognized by Swerts, Bouwhuis & Collier (1994). They found that a difference of 100 ms in the timing of the fall led to significantly different judgements about the "finality" of the contour (cf. their experiment 3). The early fall started 20 ms before the onset of the vowel in their stimuli, and the difference between the two types of fall is similar to that illustrated in Fig. 1.

In the autosegmental description of Dutch intonation by Gussenhoven (1988, 1991) and Van den Berg *et al.* (1992), the difference between contour (a) and contour (b) is interpreted as categorical. Contour (a) is taken to consist of %L H* H*L L%, as in Pierrehumbert (1980). (In Gussenhoven's analysis H* is derived from a prefinal H*L by an L-deletion rule, but this is not relevant to the issue discussed here.) Contour (b) is interpreted as having a downstepped fall on the second syllable. The contour is composed of the same sequence of H*L's, but now a morpheme "downstep" has been added to the contour. This has three effects. First, the L of the prefinal H*L is deleted, as in contour (a). Second, the target of the second H* is lowered by a downstep factor, creating !H*. Third, the target of the preceding H* spreads rightwards, creating a plateau which starts at the first accented syllable and continues until just before the second accented syllable. The falling movement at the end is thus formed by the trajectory from the target of the first, non-downstepped H* via that of the downstepped H* to that of the following L.

The phonological, discrete nature of the distinction illustrated in Fig. 1 is therefore not controversial. Indeed, there would appear to be a clear meaning difference between the two contours, although it is difficult to pin it down to particular labels. The downstepped contour sounds more as if it were meant as a definitive contribution to the discourse, and does not seem intended to draw the listener into a further discussion or evaluation of that contribution.

The problem encountered in the speech synthesis was that changes in the segmental composition of the accented syllable at times caused a downstepped fall to sound like a non-downstepped fall and vice versa, for some value of STARTIME. In such cases, the error could be rectified by adjustment of the value for STARTIME. This suggests that the timing of pitch targets is sensitive to the segmental composition of the syllable, and, moreover, that this variation is perceptually relevant.

1.2. *Phonological units* vs. *P-centres*

If the context-dependence of alignment points is perceptually relevant, as suggested by the informal observation described above, the question arises as to how this alignment point should be implemented. Is it necessary to refer to some phonological element in the syllable, i.e., the rhyme, the onset, or the nucleus? Is it necessary to refer to some articulatory element, i.e., the voiced section around the accented vowel? Or should the alignment point be equated with the perceptual centre or P-centre (cf. Ohala & Kawasaki, 1984)? Experiments on tempo perception (Pompino-Marschall, 1991) have shown that the temporal spacing of P-centres at least partly determines subjective speech tempo. If speech tempo is determined by P-centres, it should not be surprising to find that f_0-targets are aligned with these perceptual events, rather than with the segments that make up the syllables. Of course, if this were to be the case, it would not be so easy to incorporate segment-dependent variation of the alignment point in the synthesis program, because the relation between alignments and segmental make-up would be indirect.

The experiment was conducted in order to find answers to some of these questions. One of the variables selected to test was the number of consonants in the onset, within the limits imposed by the conditions on well-formed syllable onsets. Within the onset, the number of voiced consonants was varied, so that we could see if the voicing of onset consonants has an effect on the alignment point. Thirdly, the voicing of the coda consonant was varied. The locations of the P-centres were included as a fourth predictor. As argued above, the P-centre is an intuitively appealing anchor for intonational pitch movements. This concept, defined as the psychological moment of occurrence of a syllable, has been modelled in a number of ways. In some of these models (Marcus, 1981; Cooper, Whalen & Fowler, 1988), the duration of the initial consonant(s) and the syllable's rhyme or coda are the determining parameters, while other models are based on a more global spectral distribution of the energy in the syllable (Howell, 1988) or are even psycho-acoustic in nature (Pompino-Marschall, 1989, 1991). The latter model is based on the time course of the loudness within each of a number of critical bands. This model was opted for (specifically, for the version described in Pompino-Marschall, 1991), as it incorporates a large number of features of the other models, such as onset and offset events.[1] Also, it includes to some extent a number of other predictors considered here, and is reported to account fairly accurately for empirical P-centre data as obtained under different operationalisations.

[1] The 1991 version of the algorithm for the calculation of P-centres is the same as that described in Pompino-Marschall (1989), except that the weights used for the partial offset events have been reversed, after the application of a preweighting factor of 0.5.

TABLE I. Duration (in msec) of the onset consonants in the syllable *X* in the synthetic utterance *Maar ARNHEM ligt in X-land*

Onset	Consonantal duration (ms)		
	C	C	C
str	80	95	70
st		80	95
t			95
xr		50	70
l			105

2. The experiment

2.1. *Stimuli*

In order to test the hypothesis that the segmental composition of the accented syllable influences the timing of the falling pitch movement, the sentence *Maar* ARNHEM *ligt in X-land* ("But Arnhem lies in X-land") was synthesized with a number of different segmental compositions of the syllable represented by X. The onset was varied from three to zero consonants, for a total of six onset types: /str/, /st/, /t/, /xr/, /l/ and Ø.[2] Table I gives the durations of the onsets having one or more consonants. Each onset was combined with either /m/ or /p/ in the coda; the vowel was always /oː/. The rhyme had a constant duration of 215 ms (135 ms for /oː/ and 80 ms for the /p/ or /m/). (There are no substantial duration differences in Dutch between vowels before nasals and vowels before obstruents; Dutch does not contrast voiced and voiceless obstruents in the coda.)

The P-centres for these synthetic syllables as calculated with the algorithm in Pompino-Marschall (1991) are given in Table II. Inspection of the values shows that the P-centre is sensitive to the onset and offset conditions. Thus, a voiced coda corresponds with later P-centres, while longer onsets tend to correspond with earlier P-centres (the latter tendency is more easily seen in the "from syllable offset" values).

The waveforms were generated by the allophone-to-speech synthesizer RIAS (Gussenhoven & Rietveld, 1992) with a bandwidth of 5 kHz. Each of the ($6 \times 2 =$) 12 sentences was provided with an intonation contour that had a rising movement of 100 ms in the first accented syllable (*ARN*-), and a high level stretch until a falling movement of 100 ms at the end of the contour. The timing of this fall (not its slope) was varied in ten steps of 20 ms, such that in the earliest version (step 1) the fall ended 10 ms before the vowel onset, while in the latest version (step 10) it ended 35 ms after the end of the vowel. This is shown in Fig. 2. Perceptually, step 1 was clearly an instance of the downstepped contour, and step 10 was clearly a token of the non-downstepped one. This procedure resulted in ($12 \times 10 =$) 120 stimuli.

[2] Dutch [r] may be alveolar or uvular, and [x] varies from velar to uvular. The auditory impression of the [r] is that of a uvular variant. The quality of [x] is velar, rather than uvular in this synthesis program.

TABLE II. P-centres as calculated using the algorithm of Pompino-Marschall (1991), expressed in ms from syllable onset and syllable offset

	From syllable onset	From syllable offset
stroːm	116.8	343.2
stroːp	73.2	386.8
stoːm	52.6	337.4
stoːp	32.8	357.2
toːm	55.0	255.0
toːp	35.2	274.8
xroːm	158.8	176.2
xroːp	117.3	217.7
loːm	72.1	247.9
loːp	52.9	267.1
oːm	14.8	200.2
oːp	16.0	199.0

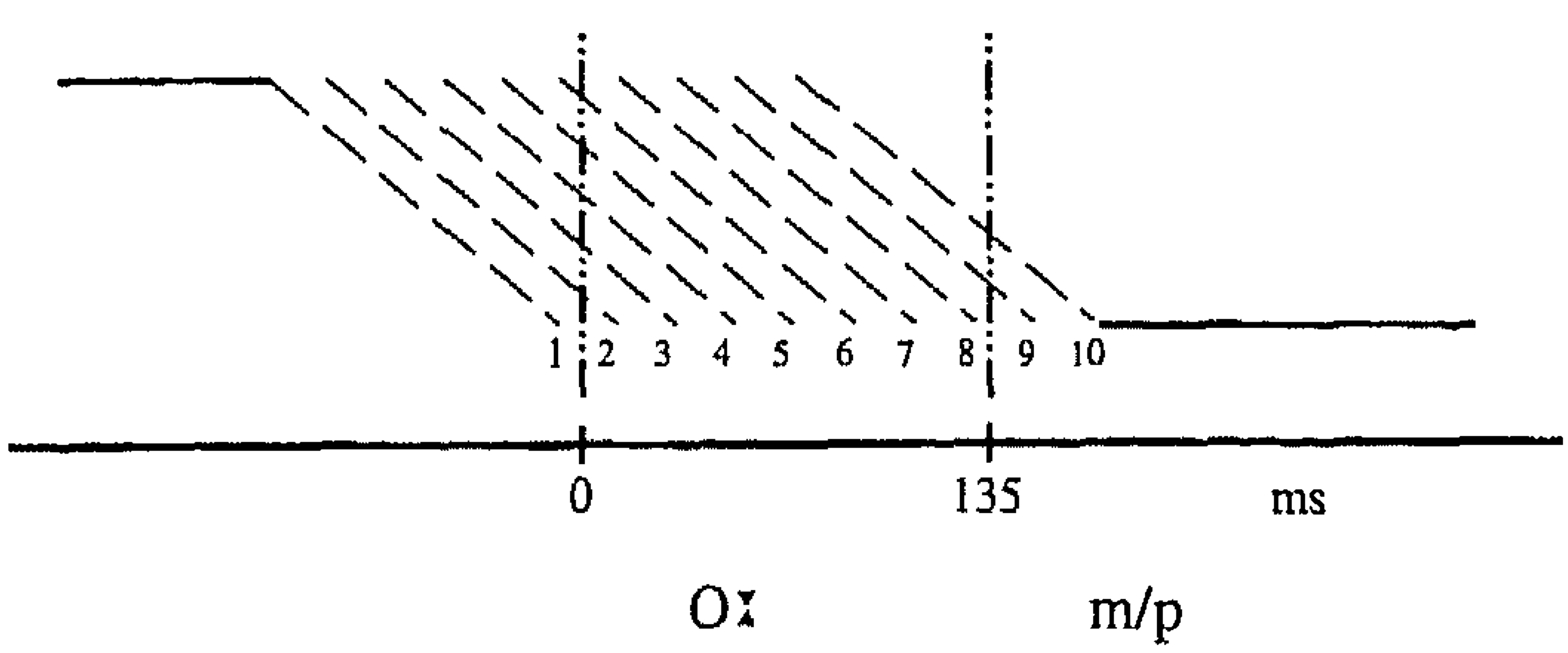

Figure 2. Locations of the F_0-falls in the stimuli: 10 steps of 20 ms.

2.2. *Subjects and procedure*

To avoid order effects, four test tapes were produced, each with a different randomized order of the stimuli. Each test version was preceded by a series of 10 minimal pairs, illustrating the contrast involved. These were presented to the subjects as a familiarization series in which the two patterns were described as "quiet, low-pitched", and "more emphatic, high-pitched", and identified as "Pattern I" and "Pattern II", respectively. The familiarization portion also included a series of six trial stimuli. The subjects received feedback and could ask questions after responding to the trial stimuli. In the test proper, two minimal pairs were included after every set of twenty stimuli, to reinforce the subjects' ability to identify the contours. Each stimulus occurred once in the test, with inter-stimulus intervals of 5 s.

Each of the four test tapes was presented via loudspeakers to one of four subgroups of native Dutch-speaking subjects. The 27 subjects were roughly equally distributed over the four subgroups. They were recruited from the student population of the University of Nijmegen and received a small fee for participating in the test. Judgements were recorded by circling a "I" or a "II" on the score sheets.

3. Results

The responses of the 27 subjects were pooled, and proportions of "downstep" responses were calculated. Following classical psychophysical procedures (Sixtl, 1967), these proportions were subsequently transformed to values of the probit-function, in order to estimate the Point of Subjective Equality (PSE). The goodness-of-fit of all transforms was assessed with the help of a χ^2-test. None of these tests was significant. The χ^2-values ranged from 4.38 ("loop") to 12.52 ("toop"), with the critical value of χ^2 being 14.07 ($p = 0.05$, $df = 7$). It is concluded, therefore, that these approximations fit the data quite well.

The PSE-values of all 12 words were calculated by $-a/b$, in which a is the intercept and b the regression coefficient of the linear regression equation $Z_i = a + bS_i$, with S_i being the rank of the stimulus, counting from the earliest stimulus, and Z_i the probit-transformed p-value. In Fig. 3, the PSE-values for each of the 12 syllable types are plotted. The vertical lines at the top of the figure represent the ten timings of the end of the falling movement, each of those steps representing a 20 ms interval.

Fig. 3 shows, for instance, that the voiceless coda of the accented syllable induces an earlier PSE in the majority of the cases. Thus, when the fall in [loːp] ends some 50 ms after the vowel onset (corresponding with step 4, which falls to the right of the PSE for "loop", marked as "O" in the bottom bar of Fig. 3), subjects perceive a non-downstepped contour. The same physical fall, however, leads to the perception of a downstepped contour if [p] is replaced with [m], as is clear from the fact that it falls to the left of the • in the same bar. (The pair

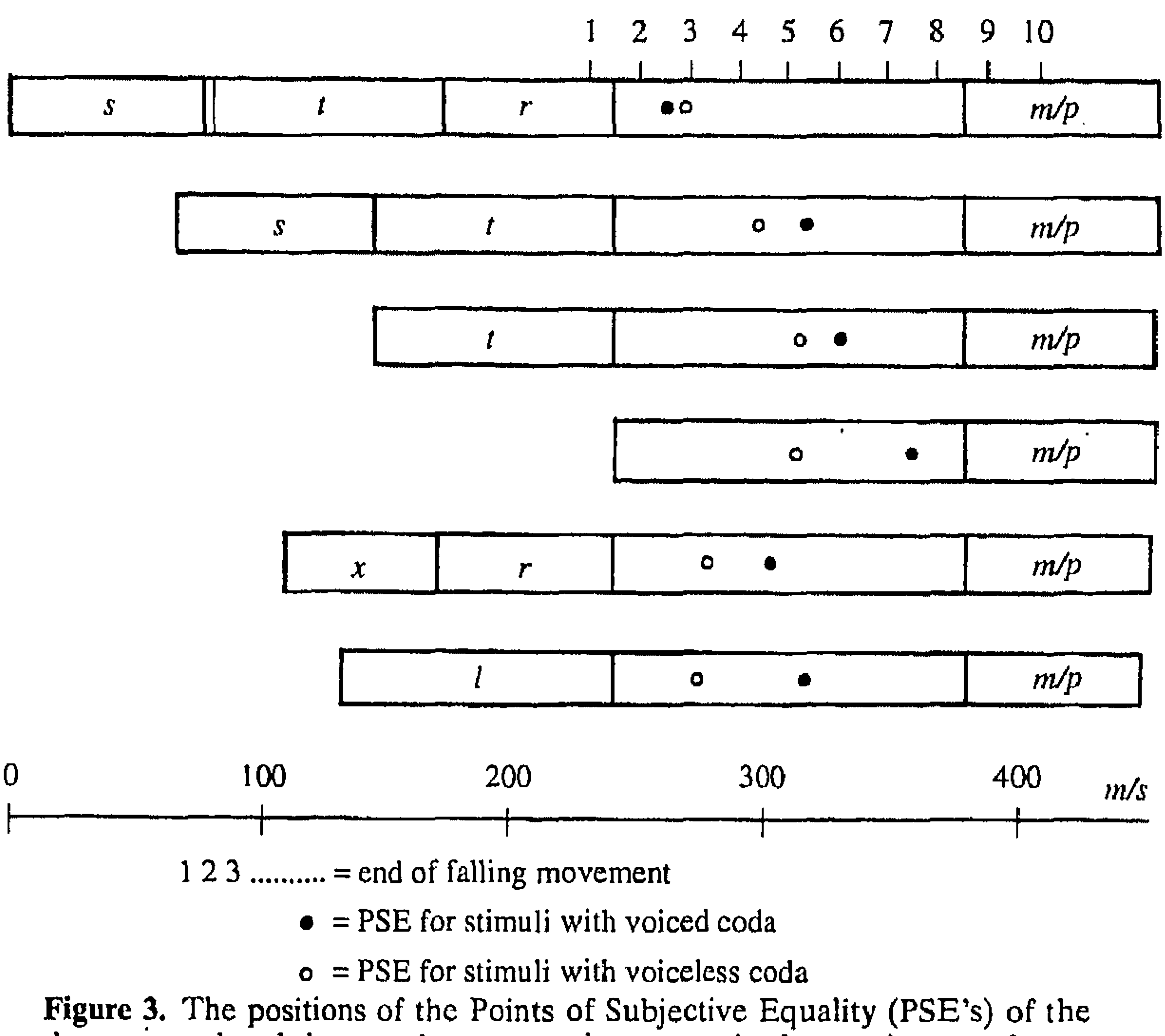

Figure 3. The positions of the Points of Subjective Equality (PSE's) of the downstepped and the non-downstepped contours in the continuum of ten timings of the endpoints of the falls, for the 12 stimulus syllables.

TABLE III. Multiple R's with explained variation in PSEs by three experimental variables

	Multiple R	Cumulative Multiple R^2
ONSETDUR	0.751	0.564
CODAVOICE	0.852	0.726
VOICEDONSETDUR	0.942	0.888

/stroːm/—/stroːp/ shows an exceptional pattern.) The mean position of the PSE for syllables with voiced codas was 5.17 and for syllables with voiceless codas 4.09, which represents a difference of 22 ms. The largest difference observed was that between /stroːp/ and /oːm/, which amounted to 109 ms.

In order to determine the factors that may underlie the variation in the location of the PSEs, a multiple regression analysis was carried out, with the following predictors:

1. The categorical variable ± voiced coda (CODAVOICE).
2. The duration of the onset (ONSETDUR).
3. The duration of the voiced part of the onset (VOICEDONSETDUR).
4. The location of the P-centre (P CENTRE).

The procedure stepwise yielded three significant predictors and a Multiple R of 0.942. The predictors are ONSETDUR ($\beta = -0.60$), CODAVOICE ($\beta = 0.40$), and VOICEDONSETDUR ($= -0.43$). All three predictors are significant at the 1% level. The minus signs of the regression weights mean that longer durations of the onset correspond to earlier positions of the PSE. A voiced coda results in a movement to the right of the PSE. Table III gives R and R^2 . As can be seen, the three significant factors together explain 89% of the variation. The P-centre did not meet the entrance requirement of the regression procedure. It is significantly correlated with the PSE, but does not have sufficient independent power to improve the prediction of the other predictors. Although the P-centre does contain elements that are relevant for the prediction of the PSE, as is clear from an inspection of Table II, the particular way they have been combined in the P-centre does not yield a better prediction than is provided by the isolated variables in the regression equation. Table IV gives the product-moment correlation-coefficients. Obviously, the correlations of CODAVOICE with ONSETDUR and VOICEDONSETDUR are 0.0, since in the stimuli the binary variable CODAVOICE was crossed with the onsets. The correlations of PCENTRE with CODAVOICE, ONSETDUR, and VOICEDONSETDUR, of which only the last reaches significance, reflect the P-centre's sensitivity to these variables.

4. Discussion

The perception experiment, which involved asking subjects to assign a set of contours with various alignments of the final fall to either a downstepped contour (early fall) or non-downstepped contour (late fall), revealed that the Point of Subjective Equality (PSE) was sensitive to the segmental composition of the

TABLE IV. Product-moment correlation coefficients between variables in the regression analysis and PSE's ($n = 12$)

	PSE	CODAVOICE	ONSETDUR	VOICEDONSETDUR
PSE	—			
CODAVOICE	0.403	—		
ONSETDUR	−0.751**	0	—	
VOICEDONSETDUR	−0.638	0	0.349	—
PCENTRE	−0.519	0.278	0.518	0.608*

* = $p < 0.05$ two-tailed.
** = $p < 0.01$ two-tailed.

accented syllable. This result is interpreted in terms of the alignment of the pitch target for the accented (starred) tone of the pitch accent. The alignment point is not a fixed location in the syllable, but depends on its segmental structure. This dependence cannot be explained by appealing to the notion of P-centre, as defined on the basis of the procedure in Pompino-Marschall (1991). Rather, the alignment point is a function of the length of the onset, and the availability of voiced consonants in the onset and in the coda. This result is explained primarily in terms of a dependence of the timing of the target of H* tones on the availability of sonorant segments in the syllable: the presence of sonorants in the onset moves the target to the left, while a voiced coda pulls that target to the right. The phonetic implementation rules apparently place targets so as to facilitate a comfortable realization of the pitch features concerned. However, it is interesting that the effect of the onset is not purely due to the presence of voiced segments, but is due to the presence of voiceless segments as well, as shown by the fact that the first variable in the multiple regression equation is the duration of the onset *per se*. This indicates that the implementation rules are sensitive to the syllable domain, and place the target with reference to its beginning. With respect to this latter point, the possibility must be left open that here too there exists a centering tendency in the placement of the target of the starred tone within the syllable. Further experimentation in which the duration of the coda is included as an experimental variable will have to decide this issue.

These results tally with the findings reported for English in van Santen & Hirschberg (1994). In a large corpus of utterances with accent-lending pitch peaks produced by a single speaker, they found that the peak time as measured from the vowel onset is later when the coda is all-sonorant than when it contains a non-sonorant consonant. This corresponds with the finding that a voiced coda pulls the target rightwards in Dutch. Second, van Santen & Hirschberg found an all-sonorant onset causes the peak time to be shifted leftwards relative to other onsets, at least when peak time is measured from the syllable onset. When measured from the vowel onset, however, an all-sonorant onset causes the peak time to shift to the right. Both these latter effects are due to the fact that the non-sonorant onset condition (which includes onsets that are partially sonorant, like [pl-, sm-]) is represented by longer onset durations in their corpus. That is, their finding is the counterpart of this study's finding that the longer the onset, the further to the left the target will move, whether measured from syllable onset or vowel onset.

Somewhat surprisingly, the P-centre, in spite of its partial dependence on the length of the onset in the model used, does not explain any additional variation in the data. Although there are different operationalisations of this concept, this would appear to represent the best-informed model, and this result should therefore be taken seriously. A caveat is in order, however, because of the synthetic stimuli used in the experiment. Replication of the finding with natural speech would be necessary before the P-centre can be definitively ruled out as an anchor point for pitch features.

This experiment has shown that the variability in the alignment point has auditory consequences, and as such should have consequences for the synthesis of Dutch accented syllables. So far, the algorithm for aligning the target for T* (STARTIME) has been solely based on the duration of the vowel in the accented syllable. In view of the results of the experiment, however, the algorithm will have to be made sensitive to other portions of the syllable. As a first approximation, a new algorithm will cause each consonant in the onset to move the alignment point leftwards by 15 ms. In addition, a voiced segment in the onset will move it leftwards by another 15 ms, while a sonorant in the coda will move it rightwards by the same amount.

We thank Jacques Terken for providing information on the timing of 'A' in the IPO-description, Julie Vonwiller for advice and information about the P-centre, and Bernd Pompino-Marschall for his invaluable assistance with the determination of the P-centres of our experimental syllables. We are very grateful for the comments by Janet Fletcher, Janet Pierrehumbert, and an anonymous reviewer, which led to considerable improvements in the exposition. We remain responsible for any errors.

References

Berg, R. van den, Gussenhoven, C. & Rietveld, T. (1992) Downstep in Dutch: implications for a model. In *Laboratory Phonology II* (G. J. Docherty & D. R. Ladd, editors), pp. 335–359. Cambridge: Cambridge University Press.

Caspers, J. & van Heuven, V. J. (1993) Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall, *Phonetica,* **50**, 161–171.

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

Collier, R. & 't Hart, J. (1981) *Cursus Nederlandse Intonatie.* Leuven: Acco.

Cooper, A. M., Whalen, D. H. & Fowler, C. A. (1988) The syllable's rhyme affects its P-center as a unit, *Journal of Phonetics,* **16**, 231–241.

Gussenhoven, C. (1988) Adequacy in intonation analysis: The case of Dutch. In *Autosegmental studies on pitch accent* (H. van der Hulst & N. Smith, editors), pp. 95–121. Dordrecht-Cinnaminson: Foris Publications.

Gussenhoven, C. (1991) Tone segments in the intonation of Dutch. In *The Berkeley Conference on Dutch Linguistics 1989* (T. F. Shannon & J. P. Snapper, editors), pp. 139–155. Lanham (MD): University Press of America.

Gussenhoven, C. & Rietveld, T. (1992) A target-interpolation model for the intonation of Dutch. In *Proceedings of the International Conference on Spoken Language Processing 92*, University of Alberta, Canada, 1235–1238.

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

House, J. (1989) Syllable structure constraints on F_0 timing. Poster presentation, Second Conference on Laboratory Phonology, Edinburgh.

Howell, P. (1988) Prediction of P-centre location from the distribution of energy in the amplitude envelope: I and II, *Perception and Psychophysics,* **43**, 90–93 & 99.

Kohler, J. J. (1990) Macro and micro F_0 in the synthesis of intonation. In *Papers in Laboratory Phonology I* (J. J. Kingston & M. E. Beckman, editors), pp. 115–138. Cambridge: Cambridge University Press.

Marcus, S. M. (1991) Acoustic determinants of Perceptual Center (P-center) location. *Perception and Psychophysics*, **30**, 247–236.
Ohala, J. J. & Kawasaki, H. (1984) Prosodic phonology and phonetics, *Phonology Yearbook*, **1**, 113–127.
Pierrehumbert J. B. (1980) *The phonology and phonetics of English intonation.* MIT Thesis. Published in 1990 by Garland Press, New York.
Pompino-Marschall, B. (1989) On the psychoacoustic nature of the P-centre phenomenon, *Journal of Phonetics*, **17**, 175–192.
Pompino-Marschall, B. (1991) The syllable as a prosodic unit and the so-called P-centre effect. *Forschungsberichte des Instituts für Phonetik und sprachliche Kommunikation der Universität München*, **29**, 66–124.
Santen, J. P. H. van & Hirschberg, J. (1994) Segmental effects on timing and height of pitch contours. In *Proceedings of the International Conference on Spoken Language Processing 94*, 719–722.
Silverman, K. E. A. & Pierrehumbert, J. B. (1990) The timing of pre-nuclear high accents in English. In *Papers in Laboratory Phonology I* (J. Kingston & M. E. Beckman, editors), pp. 72–106. Cambridge: Cambridge University Press.
Sixtl, F. (1967) *Meßmethoden der Psychologie.* Weinheim: Julius Beltz.
Steele, J. (1986) Nuclear accent F_0 peak location: effect of rate, vowel, and number of syllables, *Journal of the Acoustical Society of America* Supplement 1, **80**; s51.
Swerts, M., Bouwhuis, D. G. & Collier, R. (1994) Melodic cues to the perceived 'finality' of utterances, *Journal of the Acoustical Society of America*, **96**(4), 2064–2075.
Zitter, A. de (1992) The perceptual salience of melodic variation: contour shape and peak height, *Journal of Phonetics*, **20**, 181–188.