Planning in speech melody: production and perception of downstep in Dutch

Vincent J. van Heuven

Universiteit Leiden

Abstract

This paper studies the planning of downsteps in Dutch enumerations of two to six items long. Is it true that the available pitch range is subdivided into smaller downsteps as the number of accented items in the enumeration is larger, and how precisely do speakers programme the stepsize? Do listeners use the size of the (first) downstep to project the length of the enumeration? The results show that the size of the first downstep (when scaled in ERB) is exactly proportional to the number of items in the enumeration – indicating a high degree of planning on the part of the speaker – but that listeners are largely insensitive to the stepsize as when asked to predict the length of the utterance.

1 Introduction

It is obvious that some form of planning takes place during speech production. One line of evidence comes from research on speech errors. Nooteboom & Cohen (1974) show that speakers program their segments some seven syllables ahead relative to the moment of speaking. At the same time they programme meaningful units some seven morphemes (or words) ahead. The estimated size of the look-ahead window is derived from the distance (measured in linguistic units, i.e. syllables or words) that intervenes between the source and the target unit involved in the anticipation type of speech error. Someone who says *to colerate quite a lot* instead of *tolerate quite a lot* has mistakenly replaced the first phoneme of *tolerate* by the first phoneme of *quite*. This implies the words *tolerate* and *quite* were simultaneously present in the speaker's mind, which amounts to a sequence of four syllables.

Planning in speech production is not restricted to segmental phonetics. There is a wealth of evidence that considerable planning occurs in prosody as well. Nooteboom & Cohen (1974) reasoned that the so-called flat hat in Dutch intonation exists by virtue of planning ahead. The flat hat is the most frequently used intonation pattern that links the last and the second-but-last accent in a sentence. The pitch rises from the low declination level to the high level on the pre-final accent, remains high until the final accent, and only then drops down to the low declination line, to be followed by a low boundary tone. It follows from this description that the speaker must know at the point in time where the pre-final rise is executed, that only one more accent is to follow before the end of the sentence. Nooteboom & Cohen (1974) show that the distance between the rise and the fall accent that are linked by a flat hat, again spans some seven syllables. Planning in intonation is also visible in the course of the declination, i.e. the imaginary line that links the low pivot points in the pitch curve that can be measured in a spoken sentence. The longer the sentence, the higher the starting pitch but the slower the rate of descent, such that the speaker always reaches the same low pitch at the end of the utterance. Note that it may not be the planning of the onset pitch per se that is at issue here. Probably, when the speaker inhales, he has a rough idea of how much material he is going to speak until the next inhalation pause – which typically coincides with a deep prosodic boundary such as the end of an utterance. This is what Liberman & Pierrehumbert (1984:220) have called 'soft' pre-planning.[1] Before longer sentences, then, the speaker will take a deeper breath than before short sentences, so it is the volume of air trapped inside the lungs that primarily determines the high onset pitch rather than some complex computational act the speaker performs on the pitch contour (which would be 'hard' pre-planning in Liberman & Pierrehumbert's terms). Whether hard or soft, the reflexes of this prosodic planning enable the listener to project the end of the sentence with substantial accuracy (Grosjean, 1983, for English; Leroy, 1984, for Dutch). My present contribution targets a similar phenomenon, the perceptual effects of which, however, have received only scant attention in the literature.

'‘t Hart, Collier & Cohen (1990) describe the intonation of Dutch as a sequence of rises and falls between a lower and an upper declination line. There are five types of rise (called 1 through 5) and an equal number of falls (called A through E). Fall E is the movement of crucial importance to this chapter. It is described as a half fall, i.e., it does not drop over the full interval between the high and low declination – nominally 6 semitones – but only covers part of the interval. Fall E seems to be used in two different functions in the intonation system of Dutch. When it is the only half fall, it is sentence final, and signals an intention on the part of the speaker expressing 'what I am saying here is actually superfluous' (Keijsper, 1985; van Heuven & Kirsner, 1999). In this function, fall E indeed ends more of less midway between the high and the low declination line, or as Nooteboom and Cohen (1984:159) say: 'Fall E […] is typical for the "street call" intonation, and often creates the impression of a musical interval, a minor third' – which would place it exactly halfway between upper en lower declination in Dutch. However, fall E may also occur recursively on successive intonation phrases. In Nooteboom & Cohen (1984:161) we find seven grammatical stylised pitch contours on the sentence *Wij proberen de SPRAAK te beGRIJpen en te beHEERsen* 'We attempt the speech to understand and to control'.[2] The seventh contour shows a full rise on the accent on *SPRAAK*, and falls E on the syllables *GRIJ* and *HEER*. Since the two incomplete falls together span the distance between upper and lower declination, the stylised contour has been drawn such that each fall E spans half the distance. This pattern is the shortest exemplar of the 'terrace pattern' ('t Hart et al., 1990:166). Now, one might wonder what would happen if the utterance comprises a larger number of intonation phrases, say three, four or five, each of which would end in a partial fall E.[3] As a first approximation we would expect the speaker to divide the range between upper and lower declination into as many equal-sized intervals as are needed to make the required number of steps down. This situation typically arises in extended enumerations of the type *Ik wil een salade met MANgo, DRUIven, AARDbeien, meLOEN, DAdels en BRAmen* 'I want a salad with mango, grapes, strawberries, melon, dates and raspberries'. There would be a full rise on *MAN*, and incomplete falls E on each of the following items in the enumeration. Three incomplete falls would then require each one-third of the nominal 6-semitone span between upper and lower declination, i.e. 2 semitones (st). By the same token, a five-item enumeration would be realised as a full rise followed by four one-quarter falls E, and so on for even longer lists. If we find this specific behaviour on the part of the speaker, this would be a convincing case of so-called 'hard' pre-planning in intonation. It would indicate that the speaker knows how many items after the first item in his enumeration are to follow until the end, and then subdivides the available pitch range into the required

[1] Soft preplanning is tantamount to behavioural common sense, as opposed to 'hard' preplanning, which would involve right-to-left computation of the contour (van den Berg, Gussenhoven & Rietveld, 1992: 354-355).

[2] Throughout this chapter small caps denote pitch-accented syllables.

[3] The fall on the very last IP in a sequence of IPs is transcribed as A rather than E by 't Hart et al. (1990).

number of equal-sized steps. Moreover, it may be the case that the speaker anticipates on the length of his enumeration by making the rise on the first member of the enumeration larger as the enumeration is longer. This paper examines the speech behaviour of a sample of Dutch speakers reading aloud a set of enumerations of variable length (two to six items). We also wish to determine the communicative effects of the prosody of enumerations by asking listeners to predict how many items an enumeration will contain, when the original utterance is electronically truncated after the first downstep.

In a more recent autosegmental account of Dutch intonation (Gussenhoven, Rietveld & Terken, 1999) the terrace pattern is analysed as one possible surface realisation of a sequence of downstepped high tones (symbolised as '!H'). A downstepped high tone has a high target but less high than the immediately preceding high tone. A sequence like %L H* !H* !H* L% would then represent a terrace contour with one full rise and two incomplete falls, each spanning half the distance between upper and lower declination. Normally, however, the accents would be realised as (!)H*L configurations, so that an alternative pattern would be %L H*L !H*L !H*L L%. The two patterns, with and without the L tones, sound very much alike. It makes sense, therefore, to derive the terrace pattern from the sequence of rise-fall accents by optional deletion of the L and subsequent spreading of the high tone until the onset of the next !H, as illustrated in Figure 1.[4]

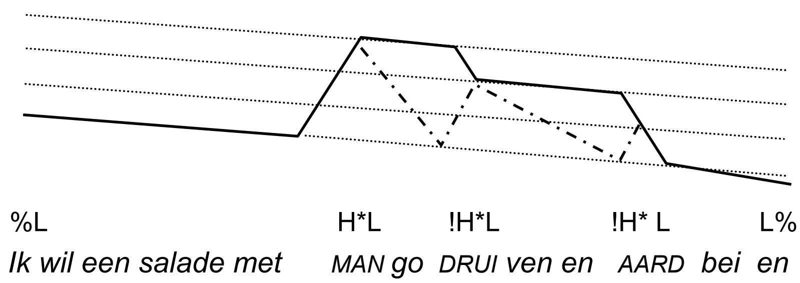

Figure 1. Underlying downstepped H*L contours (dashed) and surface terrace contours (solid).

The IPO model simply predicts that the 6-st span between upper and lower declination is bridged in two equal steps of 3 st. The autosegmental account computes the downstep values by applying a constant 'downstep ratio', such that a downstepped !H* target has a pitch (when expressed in hertz) that is a fixed percentage lower than the immediately preceding (!)H* target. Presumably the downstep ratio is smaller as the number of downsteps to be executed is larger.[5] A 3-st downstep is tantamount to a 0.8 ratio; there seems to be no principal difference here between the IPO and the autosegmental account of the phenomena. However, the autosegmental model claims that the very last downstep is considerably larger than all earlier ones – which detail is not accounted for in the IPO model.[6]

[4] Van den Berg et al. (1992: 341) suggest that the L is not deleted but squeezed so tightly onto the next !H* that it has no phonetic realisation anymore; in this latter view the H* spreading is the cause rather than the effect of the (virtual) L deletion. The autosegmental model predicts that the first downstep takes up one-third of the span between upper and lower declination (see figure 1), whilst the IPO account predicts a step down of half the span.

[5] Van den Berg et al. (1992: 354-355), however, show that (speaker-individual) constant downstep ratios across a range of enumeration lengths (two to five items) yield prediction errors that are only marginally poorer than those obtained for length-dependent downstep ratios.

[6] Liberman & Pierrehumbert (1984) include a so-called final downstep constant to account for this phenomenon in English. The magnitude of the lowering constant seems to be a language-specific parameter, and may in fact be equal to zero.

The standard version of the IPO grammar of Dutch intonation assumes that the upper and lower declination run parallel, predicting that [+full] accent-lending movements have the same excursion size irrespective of their position in the utterance. Although the onset pitch of a sentence is higher as it is longer, the excursion size of the accents remains unaffected. We cannot exclude the possibility that sentences with enumerations have a higher first accent as the enumeration is longer. If this is so, then the size of the first accent-lending rise might be a perceptual cue to the length of the enumeration (see also note 5). A second(ary) cue would then be the size of the first downstep, which would either be manifested as the interval of the fall after the first terrace, or the interval between the H*L and the first !H*L accents. The smaller the pitch interval between these two accents, the longer the enumeration should be.

2 Production experiment

2.1 Methods

Nine adult native speakers of Dutch (five male, four female) read the following ten sentences in different random orders each:

A2. Ik wil een salade met MANgo en DRUIven.
A3. Ik wil een salade met MANgo, DRUIven en AARDbeien.
A4. Ik wil een salade met MANgo, DRUIven, AARDbeien en meLOEN.
A5. Ik wil een salade met MANgo, DRUIven, AARDbeien, meLOEN en DAdels.
A6. Ik wil een salade met MANgo, DRUIven, AARDbeien, meLOEN, DAdels en BRAmen.
‘I want a salad with mango, grapes, strawberries, melon, dates and raspberries’

B2. Op mijn lijstje staan BLOEM en BOter.
B3. Op mijn lijstje staan BLOEM, BOter en EIeren.
B4. Op mijn lijstje staan BLOEM, BOter, EIeren en ROOM.
B5. Op mijn lijstje staan BLOEM, BOter, EIeren, ROOM en MELK.
B6. Op mijn lijstje staan BLOEM, BOter, EIeren, ROOM, MELK en roZIJnen.
‘On my list are flour, butter, eggs, cream, milk and raisins’

Before reading the stimuli speakers heard and repeated exemplars of a sentence with a (lexically different) three-item enumeration, with stylized resynthesized terrace contours in order to improve the odds that the speakers would then continue to use this intonation pattern when reading the actual stimuli.

Recordings were made in a sound-insulated booth through a Sennheiser MKH 416 uni-directional condenser microphone onto computer disk (16 bit, 16 kHz). Fundamental frequency curves were measured using the autocorrelation method implemented in Praat (Boersma & Weenink, 1996), interactively corrected when the algorithm had made an error, and psychophysically scaled in ERB units (see Nooteboom, 1997, and references therein).

2.2 Results

As a first approximation I present a rather abstract representation of the pitch contours measured in the materials. For each utterance, Figure 2 displays the F0 value (in ERB) for the onset pitch (representing the low boundary tone %L), the low (L) immediately preceding the accent-lending rise on the first accented syllable in the enumeration, and the terminal F0 (L%). Between L and L% there is an array of measurement point representing the F0 maxima measured on the accented syllables in the successive items of the enumerations. The maxima are either true peak values (in the case of rise-fall accents) or the beginnings of rather level tones (when a terrace-type of contour was realised). The measurements are collapsed separately over the five male and four female speakers, and over the two lexically different series of enumerations, but are broken down by enumerations of different lengths. Note that although the contours all share the same beginning of the utterance (the %L and L measurement points), they are lined up on the *end* of the contour, such that, for instance, the last downstepped H* ('H-0') accent peaks are always at the same position along the time axis. This alignment of the measurement points affords the clearest view on what is going on in the data. The X-axis represents a continuous time axis only for the longest contours (with length = 6). For all other contours, the time axis is discontinuous after the low onset of the accent-lending rise on the first item in the enumeration (indicated by in interruption of the lines for length 2 through 5 in figure 2). Note that the X-axis in this figure abstracts away from physical duration: all pitch pivot points are drawn at equal 'time' intervals.

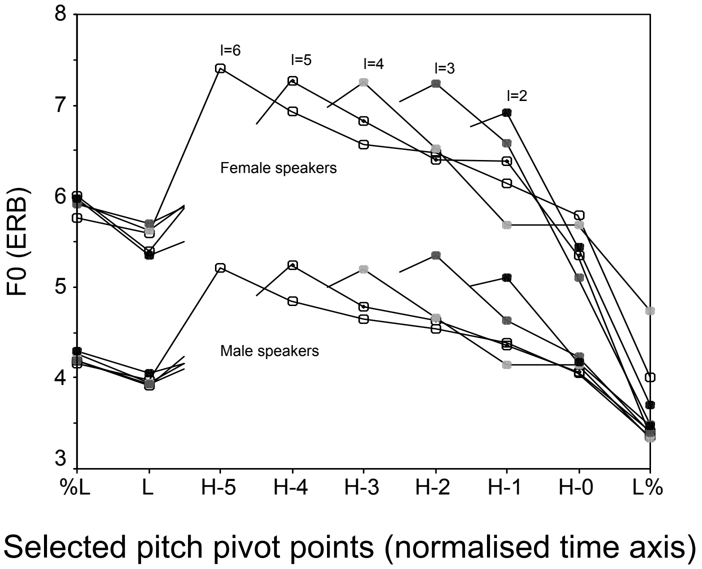

Figure 2. Fundamental frequency (F0 in ERB) of selected pivot points (see text) in contours produced by five male and four female Dutch speakers on enumerations of two to six items long. The time axis is normalized by inter-pivot distance, and discontinuous for all contours except for those with length = 6.

Three-way analyses of variance were done on the measured frequencies (in ERB) for each of the selected pivot points, with length of enumeration and sex of speaker as fixed factors and with the two lexical sentence types as a random factor. Table 1 lists the results for the effect of length of enumeration and for the interaction between length and sex of speaker. Of course, the effect of sex on the measured F0 values was always highly significant; given that the repetition rate of female voices is typically twice that of male voices, this is a predictable if not trivial effect, which we have not included in Table 1. The ANOVA results confirm the visual impression from Figure 2. The mean pitch values at the first two and the last two pivot points do not vary as a function of the length of the enumeration. The stretch of low declination in the precursor phrase is the same for all enumerations, and so is the pitch of the sentence-final downstepped H* accent (H_0) and the terminal pitch of the sentence (L%).

Counter to 't Hart et al. (1990), then, longer utterances do not start on a higher pitch than shorter utterances, at least not in this type of material. It is also apparent from Figure 2 that the F0 peak at the rise-fall accent on the first item of the enumeration is roughly constant across the entire range of enumerations. A one-way ANOVA on the peak F0 of the first accent shows no effect of the number of items to follow in the enumeration, $F(4,82)<1$. By and large, then, it seems that the first H* and the last downstepped !H* targets are constants, irrespective of the number of downsteps that should be executed in between these two. It follows from this characterisation that the mean downstep size is smaller as the number of downsteps in the enumeration is larger. However, for an enumeration of a specific length the downsteps are equal, at least when F0 is expressed in ERB. That is, the downstep size is a linear function of the number of items in the enumeration. There seems to be no need, then, for a special lowering constant on the last downstep, as was advocated for English (Liberman & Pierrehumbert, 1984) and Mexican Spanish (Prieto, Shih & Nibert, 1996).

Table 1. Results of three-way ANOVA. Effect of length of enumeration and interaction of length × sex on nine selected pitch pivot points in intonation contours.

Pivot point	Length of enumeration				Length × sex			
	df_1=df_2	F	p	sign.	df_1=df_2	F	p	sign.
%L	4	3.9	.108		4	.4	.799	
L	4	.6	.684		4	2.0	.254	
H_{-5}								
H_{-4}	1	20.2	.139		1	1.9	.404	
H_{-3}	2	47.2	.023	*	2	.2	.822	
H_{-2}	3	17.7	.021	*	3	.2	.896	
H_{-1}	4	7.4	.039	*	4	2.2	.235	
H_0	4	.5	.767		4	3.9	.108	
L%	4	9.4	.026	*	4	6.1	.054	

Note: No effect or interaction can be measured at pivot point H_{-5}, since length of enumeration is a constant there.

As a final point in the analysis of the production data, let us consider the effect of length of the enumeration, i.e. the number of downsteps to be executed between the rise on the first item and the end of the sentence, on the size of the first downstep, i.e. the downstep on the second item in the enumeration. This is the first and the largest downstep in the sequence, and as such it potentially contains the earliest and clearest pitch cue that might allow the listener to project the length of the enumeration. Figure 3 presents the size of the first downstep as a function of the number of items in the enumeration for male and female speakers.

Figure 3 shows quite clearly that the size of the first downstep gets smaller as the number of items in the enumeration is larger. Since a constant range has to be divided into two, three, four etc. steps, depending on the length of the enumeration, we expect a reciprocal function of the type $\hat{y} = b_0 + b_1/x$ to optimally capture the relationship. Indeed the reciprocal function yielded the best fit to the data (better than logarithmic, power, exponential or linear fits). The relationship is more regular for the male speakers than for the females. Interestingly, the female downsteps tend to be larger than those of the men for the larger downsteps in shorter enumerations but asymptote to the same size across sexes for the longer enumerations.[7]

[7] The overall effect ties in with earlier observations that Dutch female speakers have larger pitch movements than their male counterparts, even when pitch is psychophysically scaled in ERB (Haan & van Heuven, 1999). Although this might also be taken as a cue that the ERB scale could be invalid for cross-sex comparison of pitch, I now believe that Dutch women genuinely have larger pitch movements than men. This belief is based on the fact that it is precisely the reciprocal function that captures the regularity in the downstep size most successfully. If the pitches had not been scaled in ERB, this result would not have been obtained.

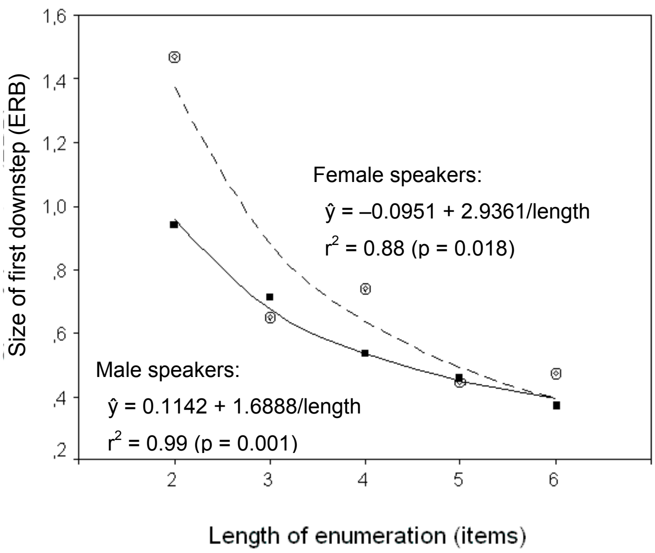

Figure 3. Size of first downstep (ERB) as a function of number of items in the enumeration for male and female speakers. Reciprocal functions have been fitted to the data points.

2.3 Conclusions

It seems that some planning mechanism has to be assumed in order to account for the results of the production study presented above. However, the speakers limit the burden of tonal planning to a minimum. For instance, they do not anticipate on longer utterances by starting off at a higher pitch nor do they execute a larger accent-lending rise on the first item in the enumeration. Rather, speakers realise all utterances, irrespective of their length, in the same register, bounded by a constant low onset pitch, a constant maximum on the first accent, a constant maximum on the last accent and a constant terminal pitch. The only planning that is brought to the task, is that the pitch range between the maximum on the first and the last accent is subdivided into a number of roughly equal-sized steps (when expressed in ERB); the step size then only depends on the number of steps that have to be walked down the terrace, i.e. on the number of items in the enumeration. So the speaker must know how many items his enumeration contains in order to select the required size of downstep.

For a perceptual follow up, we predict, then, that the only type of cue that allows the listener to project the length of the enumeration, is the size of the first downstep. So, in a gating experiment in which a larger initial portion of the utterance is made audible on successive presentations, the listener will have to wait until he has heard the portion of the utterance up to and including the second item in the enumeration; this is the point in time where the maxima of the first and second accent can be compared so that the size of the downstep can be computed. The size of the downstep should then be divided into the size of the accent-lending rise on the first item, and the listener will know how many downsteps will follow. These predictions will be tested in the next section.

3 Perception experiment

3.1 Introduction

In the perception study we targeted several potential cues that might help the listener project the length of an enumeration. As explained in section 1, the speaker might have chosen to execute a larger pitch rise on the first accent as the enumeration contains more items. Even though our speakers did not exhibit this behaviour, it is still be possible that the listener might use such a cue when it is contained in the speech stimulus. Therefore we artificially decreased the size of the first accent (i.e. the peak F0 value of the accent) relative to the natural F0 used by the speaker. Also, we know that speaking rate is faster as the sentence is longer. Consequently, when hearing a fast initial portion of an utterance, the listener will be more likely to project a long enumeration than when hearing a relatively slow onset portion. In order to be able to determine the relative contribution of the three factors identified here, i.e. (i) speaking rate of utterance, (ii) size of first accent-lending rise and (iii) size of first downstep, these factors were systematically varied in the stimulus materials.

3.2 Methods

The basic materials for this experiment were the five sentences A2 through A6 as spoken by one of the five male speakers (the present author) in the production experiment. The sentences were divided into the precursor *Ik wil een salade met* 'I want a salad with' and the enumeration part. Precursors were taken from the utterances with enumerations containing two, four and six items; these had durations of 990, 910 and 810 ms, respectively. In order to keep the experiment reasonably short, no precursors originally spoken before three and five-item enumerations were used. The three precursors, with their natural speaking rate, were cross-spliced onto each of the five enumeration portions, yielding a set of three original utterances and twelve hybrids. The declination of the precursor to the four-item enumeration was stylised and used for all the stimuli. The enumeration portions were also stylised until the end of the second item, using a straight-line approximation of the accent-lending rise H* and the following terraces separated by the first downstep. The size of the downstep was chosen to be roughly in line with the values actually found in the five natural original utterances. The largest downstep, as found for the two-item enumeration, was set at 20 Hz; the smallest downstep, found for the six-item enumeration was given a value of 10 Hz. Intermediate downsteps were interpolated at 12.5, 15 and 17,5 Hz.[8] In order not to give away lexical clues, the conjunctive *en* 'and', which regularly precedes the last item of an enumeration in Dutch, was removed using the Praat waveform editor.[9] The accent-lending rise on the first item was either given its natural value (i.e., as produced by the speaker) or 0.25 ERB less or 0.50 ERB less. When a manipulated rise was used, the terrace patterns were shifted down in frequency so as to be precisely linked to the F0 peak of the rise on the first item. All stimuli were truncated immediately after the accented syllable of the second item in the enumeration, i.e. after /ˈdrœy/ in the word *DRUIven*.

In all 45 stimuli resulted from these manipulations: 3 (precursor rates) × 3 (rises on first item) × 5 (downsteps after first item). These were presented twice in different random orders to thirteen adult native Dutch listeners, in individual interactive sessions, over headphones in a quiet room. Subjects were instructed to indicate, with forced choice, for each utterance they heard, which of the five enumerations A2 through A6 they thought had most likely just been played to them. During the experiment the five sentences were displayed on a monitor in front

[8] The size of the downsteps was manipulated in the hertz domain for reasons of convenience. Within the very restricted range of F0 we needed for the stimulus manipulations, the ERB scale and the hertz scale are practically interchangeable. Note also that no systematic differences in temporal organisation were found for the enumeration parts of the utterances. Although there may still be subtle cues remaining in the segmental or temporal make-up of the five different enumeration parts, we take the view that the perceptual results are predominantly due to the manipulation of the downstep.

[9] Strictly speaking, the resulting stimulus is ungrammatical. Nevertheless, our listeners proved perfectly capable of performing their task (see results).

of the listener, who was instructed to click on one of five radio buttons which preceded the five sentences on the screen. After each response there was a 2-second pause before the presentation of the next stimulus.

3.3 Results

An analysis of variance on the estimated length of the enumeration (expressed in number of items) with precursor rate, rise and downstep as fixed factors revealed that precursor rate [$F(2,1125)=13.1$, $p<.001$] and downstep [$F(4,1125)=53.0$, $p<.001$] yielded significant main effects. There was no effect of the size of the accent-lending rise on the first item of the enumeration, nor were any of the interactions among the three factors significant ($F<1$ in all cases). The effects of precursor rate and of downstep on the length of the enumeration as estimated by the listeners, are displayed in Figure 4.

Figure 4 shows that the mean estimated length of the enumeration went up as the speaking rate of the precursor phrase was slower. A mean length of 3.68 items was estimated for the slowest precursor, 3.79 items for the intermediate rate and the highest estimate (4.11) was found for the fastest precursor. A post-hoc test for contrast (Scheffé, $p<.05$) shows that the fastest rate differs from the two slower rates, which do not differ significantly from one another.

A post-hoc test on the effect of downstep shows that the largest downstep (20 Hz) leads to the prediction of the shortest enumeration (2.84 items). The four smaller downsteps (10, 12.5, 15 and 17.5 Hz) yield estimations of enumeration lengths between 3.97 and 4.23 items) but these downstep sizes do not differ significantly from each other. It would seem, then, that the largest downstep (20 Hz) was in fact so large that it prompted our listeners to assume that the speaker would lower his pitch all the way down to the low declination line, which of course would then be an unmistakable cue for finality, i.e. that no more items were to follow after the word *DRUIven*. When no finality cue could be picked up, the estimates of the remaining number of items in the enumeration were roughly in the middle of the possible range, i.e. enumerations between 3 and 6 items long.

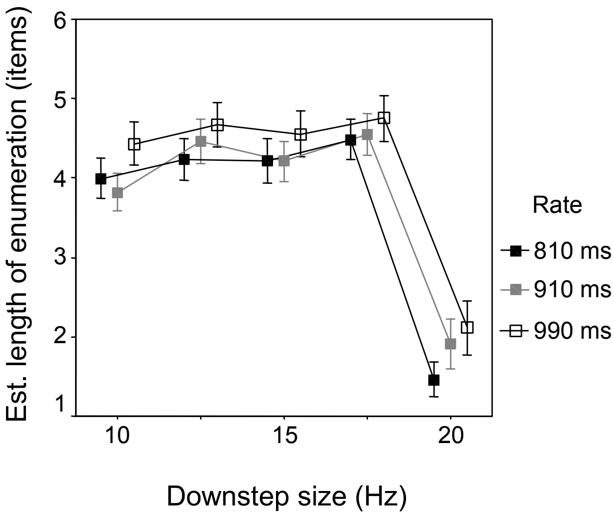

Figure 4. Estimated length of enumeration as a function of speaking rate of precursor phrase (duration in ms) and of size of downstep (Hz). Error bars represent ±2 SE.

We conclude, therefore, that neither the excursion size of the first accent-lending rise, nor the size of the first downstep provide any perceptually useful cue that allows the listener to project the length of an enumeration. The successful prediction of a two-item enumeration for the largest downstep is not a downstep cue but would rather seem to be an experimental artefact: the pitch at the moment of truncation has dropped so low that the listener takes it as a cue for finality. The only perceptually useful element of preplanning that remains in the data, is to be found in the speaking rate of the precursor phrase to the enumeration.

4 Conclusion and discussion

Our production data confirm that pre-planning plays a role in the production of speech melody. In fact, it would seem tempting to claim that some form of 'hard' pre-planning is at stake, because the details of the phonetic specification of the first downstep in the enumeration reflect a mathematical subdivision of the available pitch span. However, this may be overstating the case. The results presented in Figure 3 have been collapsed over five male and four female speakers, and there was considerable between-speaker variability in the details of the downstep sizes. So, on aggregate, the more realistic conclusion must be that the production data do not afford a stronger conclusion than that there is 'soft' pre-planning in the production of speech melody in Dutch.

The perceptual effect of varying the size of the first downstep is a simple dichotomy. Only the difference between a fairly large downstep, indicating that the speaker is aiming for a two-item enumeration, versus a whole range of smaller downsteps, cueing a longer than a two-item enumeration, is reliably picked up by the listener. Moreover, given that the speaking rate in the onset portion of the sentence provides a more successful cue to the length of the enumeration, the later cue in the first downstep would seem to function as a confirmation rather than a refinement of the early cue. More definitive evidence on the temporal distribution of the cues would have to come from more elaborate gating experiments, similar to our studies of the perceptual cues underlying the difference between statement and (declarative) question in Dutch (van Heuven & Haan 2002).

Acknowledgement

The production and perception experiments reported here were run by my students Bart van Bezooijen and Mathieu Fannée, respectively, as a course requirement for the Fall 2003 edition of the Seminar in Experimental Phonetics taught as part of the Undergraduate Linguistics Program at Universiteit Leiden.

References

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

Boersma, P., & Weenink, D. (1996). *Praat, a system for doing phonetics by computer*. Report nr. 132. Amsterdam: Institute of Phonetic Sciences, University of Amsterdam.

Grosjean, F. (1983). How long is the sentence? Prediction and prosody in the on-line processing of language. *Linguistics, 21*, 501-529.

Gussenhoven, C., Rietveld, T. & Terken, J. (1999). *ToDI: Transcription of Dutch Intonation*. Available: http://www.lands.let.kun.nl/todi .

Haan, J., & Heuven, V. J. van (1999). Male vs. female pitch range in Dutch questions. In J.J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A.C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, August 1-7, 1999* (pp. 1581-1584).

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

Heuven, V. J. van, & Haan, J. (2002). Temporal development of interrogativity cues in Dutch. In C. Gussenhoven & N. Warner (Eds.), *Papers in Laboratory Phonology VII* (pp. 61-86). Berlin: Mouton de Gruyter.

Heuven, V. J. van, & Kirsner, R. S. (1999). Interaction of grammatical form and intonation: Two experiments on Dutch imperatives. In R. Kager & R. van Bezooijen (Eds.), *Linguistics in the Netherlands 1999* (pp. 81-96). Amsterdam: John Benjamins.

Keijsper, C. E. (1984). Form and meaning of Dutch pitch contours [in Dutch]. *Forum der Letteren, 25*, 20-37, 113-126.

Leroy, L. (1984). The psychological reality of fundamental frequency declination. *Antwerp Papers in Linguistics*, *40*.

Liberman, M. & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. Oerhle (Eds.), *Language sound structure* (pp.157-233). Cambridge, MA: MIT Press.

Nooteboom, S. G. (1997). The prosody of speech: melody and rhythm. In W.J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 640-673). Oxford: Blackwell.

Nooteboom, S. G., & Cohen, A. (1974). Anticipation in speech production and its implication for perception. In A. Cohen & S.G. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 124-142). Berlin: Springer. Communication and Cybernetics; 11.

Nooteboom, S. G., & Cohen, A. (1984). *Spreken en Verstaan: Een nieuwe inleiding tot de experimentele fonetiek* (2nd, revised ed., in Dutch). Assen: Van Gorcum.

Prieto, P., Shih, C. & Nibert, H. (1996). Pitch downtrend in Spanish. *Journal of Phonetics, 24*, 445-473.