# AUTOMATIC RECOGNITION OF SENTENCE TYPE FROM PROSODY IN DUTCH

*Vincent J. van Heuven*, Judith Haan** and Jos J.A. Pacilly**
*Phonetics Laboratory, Department of Linguistics and Holland Institute of Generative Linguistics, Leiden University
Cleveringaplaats 1, PO Box 9515, 2300 RA Leiden, The Netherlands
**Department of General Linguistics and Dialectology, Centre for Language Studies, Nijmegen University
Erasmusplein 1, PO Box 9103, 6500 HD Nijmegen, The Netherlands

**ABSTRACT**

This paper investigates to what extent statements, Wh-questions, Yes/No-questions and declarative questions in Dutch can be automatically discriminated on the basis of global and local $F_0$-parameters. Global parameters were the slope and mean pitch of upper and lower trend lines that were fitted through $F_0$-curves; local parameters were onset and offset $F_0$ of a terminal question-marking pitch rise. Results indicate that women mark the interrogative status of a sentence more often and perceptually more saliently. Generally, global downtrend parameters are better predictors of sentence type than parameters of the final rise.

## 1. INTRODUCTION

Dutch, as well as a host of other languages, distinguishes at least two sentence types: statements (S) and questions (Q). Questions are further subdivided into three types:

W *Wh-questions* (also called 'information questions'), beginning with a question word (*who, where, when, what, why*, etc., and often followed by inversion of subject and finite,
Y *Yes/no-questions*, syntactically marked by inversion of subject and finite,
D *Declarative questions*, which have the same lexical items and word order as the corresponding statement.

Lexico-syntactic marking of interrogativity, therefore, is strongest in W, weaker in Y, and absent in D. All three Q-types are claimed to be prosodically marked; we expect the phonetic/prosodic interrogativity markers to be stronger in inverse proportion to the (number of) lexico-syntactic markers.

Cross-linguistically, question intonation has always been strongly associated with a local terminal rise in pitch. In part of the literature question intonation is still largely identified with this final rise [1,2]. Still it would seem more accurate to say that questions are universally marked by the presence of a 'high' element somewhere in the sentence [3]. This high pitch may manifest itself both locally, e.g. in the initial, medial or final portion of the utterance, and globally, either in the guise of a raised register[1] [4,5,6] or of the absence of $F_0$-downtrend; presence of $F_0$-downtrend is commonly observed in and across statements [7,8]. The distribution of high elements may differ between and within languages.

In Dutch, the Q-shibolet is a terminal rise, which is, in fact, the only Q-marker acknowledged in current models of Dutch intonation [9,10]. Yet, there are early claims in the literature that the Dutch question contour is hammock-shaped, i.e., it has a high beginning, a low stretch in between and an equally high ending [11,12]. It has also been claimed that Dutch questions are realized in a higher register [11,13]. Recent experiments have indeed brought to light that Dutch questions differ considerably from statements in terms of pitch range and downtrend [14]:

- $F_0$-level at the onset of the utterance (higher for Q than for S),
- degree of global $F_0$-downtrend (downwards for S vs. more level or even upward for Q),
- height of $F_0$-maxima in prominent syllables (greater in Q than in S), and
- level of the low pitches in non-prominent syllables (greater in Q than in S).

Since most of these Q-markers develop as the utterance progresses, it is important to determine whether there is a point in time in the utterance *before the onset of the terminal rise* where differences are sufficiently large to safely identify the sentence type as S, W, Y, or D. The purpose of the present research, then, is to determine to what extent sentence type in Dutch can be automatically recovered from the prosody of the utterance, at some early point in the sentence, i.e., before the presence vs. absence of a sentence-final pitch rise provides a decisive cue.

## 2. METHOD

### 2.1. Speech materials.

The basic materials are a database of 800 Dutch

1 Register is defined here as an area within a given speaker's overall pitch range, enclosed by the highest and lowest frequency within which tones of a particular utterance are realized [15].

speech utterances collected from five male and five female adult speakers of Standard Dutch. Each speaker produced two repetitions of two sentences in each of four sentence types (S, W, Y, D), either spoken in isolation or in a minimal context, i.e. in a two-sentence paragraph. In the later case the target sentence was systematically varied over the two positions.

Recordings were made onto DAT in a sound-proofed studio (48,1 kHz, 16 bit) using a Sennheiser MKH-416 condenser microphone. Subjects were asked to speak the sentences as if they were actors in a radio play.

### 2.2. Acoustic analysis

The acoustic analysis proceeded along the following steps:

- First, $F_0$ was extracted by the method of subharmonic summation [16], followed by curve smoothing over a 50-ms time window. $F_0$-values were expressed in Equivalent Rectangular Bandwidths (ERB), which is currently held to be the psychophysically most relevant $F_0$-scale in intonation languages [17].
- Next, the terminal pitch rise, if present, was segmented off from the earlier part of the target utterance by hand, using the following procedure:
  – First, two intonologists independently determined whether a final rise was pronounced by the speaker. In all cases the intonologists concurred in their judgments (apparently, speakers are quite outspoken in their terminal question markings in this type of play-acting task)
  – If a rise was found present, the end point of the rise was located at the frame in the smoothed $F_0$-curve with the highest $F_0$ within the final 250 ms of the utterance
  – The onset of the terminal rise was then defined as the latest (rightmost) local $F_0$-minimum before the end of the utterance. In a number of cases, however, no local minimum could be found; this happens when the speaker produces an accent-lending pitch rise near the end of the sentence, which is then followed by the terminal question rise without a local minimum separating the accent-lending rise from the terminal question rise – although the there was always a more or less level stretch of $F_0$ before the onset of the terminal rise[2].
- Onset $F_0$, minimum $F_0$ and maximum $F_0$ were determined for the utterance *minus the terminal rise.*
- Upper and lower declination lines (downtrend) were then fitted to the $F_0$-curve as follows:
  – A linear $F_0$-with-time regression line was fitted through the utterance (minus the terminal rise); the $F_0$-measurement points were divided into an upper and lower half, i.e. those points lying above the all-points regression line, and those lying below it, respectively.
  – Next, an upper regression line was fitted through the upper $F_0$-points, and a lower regression line through the lower points. The means at the temporal mid-point and slope coefficients of the upper and lower regression lines were the target parameters.

Figure 1 illustrates the various measurements.

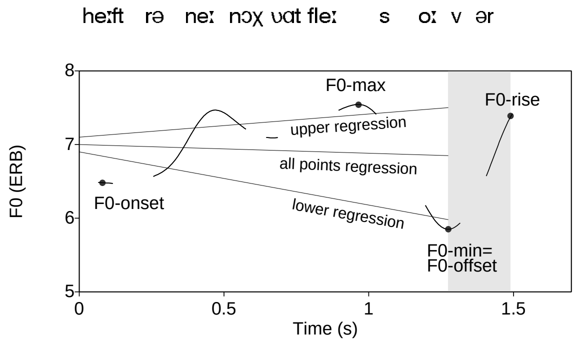


*Figure 1. Raw parameters and regression lines for an isolated Yes/No-question spoken by a male ‘Heeft Renée nog wat vlees over?’ (Does Renee have any meat left?). The terminal rise extends over the gray area.*

## 3. RESULTS

### 3.1. Contribution of local terminal pitch rise

Figure 2 presents the number of final rises detected for each of the four sentence types, broken down by sex of speaker.

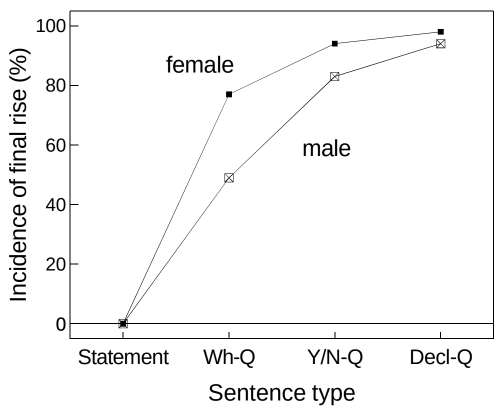


*Figure 2 Incidence of terminal rises for four sentence types broken down by sex of speaker.*

2 Precisely for this reason we have not (yet) been able to devise a reliable automatic procedure for segmenting off final rises in our material.

Final rises never occur after statements. Rises are nearly always present after declarative questions, and less often after Y/N and Wh-questions in the predicted order W<Y<D. Moreover, women clearly use more final rises than men; the difference is especially marked in Wh-questions. The greater incidence of final rises with women has been reported in the literature [18].

Figure 3 presents the onset and offset $F_0$, and implicitly excursion size, (in ERB) of those final rises that were present, broken down by question type and by sex of speaker.

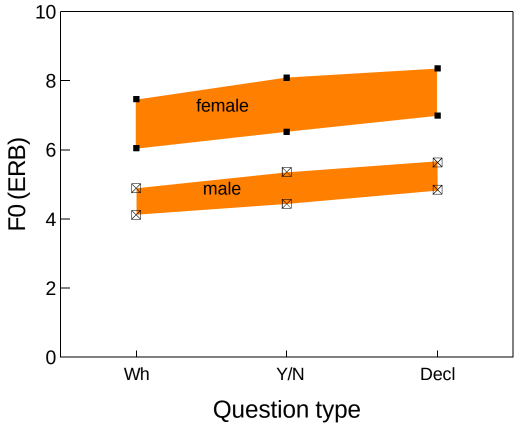


*Figure 3. Onset and offset $F_0$ of question-marking final pitch rise (ERB) for three question types broken down by sex of speaker.*

Figure 3 shows quite clearly that the excursion size of the rise is the same for the three question types. However, the onset (and offset) $F_0$ does differ, with higher values in the predicted D>Y>W order. Note that the excursion size of the female rises is roughly twice as large as that of males. Since $F_0$ is already scaled so as to optimally reflect auditory distance, this finding indicates that women mark their question rises perceptually more saliently than men.

### 3.2. Contribution of global parameters

Figure 4 summarizes the effects of sentence type on the means and slopes of the upper and lower regression lines fitted through the $F_0$-curves. These results have not been broken down for sex of speaker here; the effect of sex is as predictable from the earlier results for the terminal rise: the female register is located at a higher frequency and spans a (perceptually) wider range.

We will not comment on the effects visible in figure 4; the reader if referred to [14] instead. Here we will concentrate on the analysis of the classificatory power of the global $F_0$-parameters.

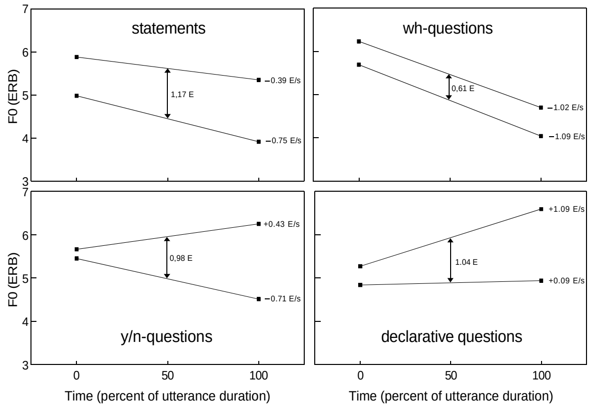


*Figure 4 Global $F_0$-parameters (upper and lower regression lines fitted through $F_0$-curve, broken down by sentence type.*

The contribution of global parameters to the automatic determination of sentence type was examined by means of a series of linear discriminant analyses (LDA) [19] on the global pitch parameters collected for the non-terminal part of the utterances, after Z-normalisation for pitch of the individual speakers in the ERB-domain; the Z-transformation abstracts from both differences in mean $F_0$ among speakers as well as from the wider register used by the women. The performance of the LDA is shown table 1.

*Table 1 Classification of all four sentence types by Linear Discriminant Analysis from global parameters only (i.e. excluding properties of final rise). Correct classifications along the main diagonal.*

| actual sentence type | predicted sentence type | | | |
|---|---|---|---|---|
| | S | W | Y | D |
| S | 82 | 9 | 5 | 4 |
| W | 8 | 75 | 17 | 0 |
| Y | 12 | 12 | 53 | 23 |
| D | 11 | 4 | 18 | 67 |

These results show that there is little confusion between S and Q utterances. Even without taking the final rise into account, S is identified correctly in 82% and Q in 90%. The internal classification of Q-subtypes is poorer (between 53 and 75% correct, but always well above chance. Notice that Y and D questions resemble each other more (i.e., are more often confused) than any of the other sentence types. Given that final rises do not differ among the question types (only their incidence does, see above), we will finally attempt an LDA on global $F_0$-parameters for only those utterance that do not end in a rise. Table 2 presents the results:

*Table 2 As table 1, but only for utterances that do not end in a question-marking pitch rise.*

| actual sentence type | predicted sentence type | | | |
|---|---|---|---|---|
| | S | W | Y | D |
| S | 89 | 5 | 3 | 3 |
| W | 4 | 80 | 16 | 0 |
| Y | 4 | 5 | 73 | 18 |
| D | 33 | 0 | 11 | 56 |

These results demonatrate that global parameters differentiate rather well among the various sentence types, even if they were never pronounced with a terminal rise. The confusion between statements and Wh-question remains substantial, however.

## 4. CONCLUSION

A terminal rise is a certain predictor of the question status of an utterance, at least in the type of materials used in this study. Absence of a rise is not a reliable predictor of the statement type: many questions, especially Wh and Y/N-questions, lack the final rise.

Although the data analysis is still in the initial stages, these intermediate results are promising, showing that (automatic) recognition of sentence type is feasible, certainly in terms of a binary decision (statement vs. question) and somewhat less, though much better than chance, in terms of the more refined four-category decision (statement, W-question, Y-question, D-question). Finally, the confusion rate for the three Q-types with S shows the order W<Y<D in the performance of the LDA. This supports our prediction that the prosodic marking of interrogativity will be stronger as the number of lexico-syntactic markers is smaller. Women, finally, mark their questions more often and perceptually more saliently than men.

Practical application of these results is to be sought in the field of automatic speech understanding and dialogue systems. Naturally, it is of considerabe importance that such systems recognize statements from questions. Prosody then provides useful information on the speaker's intention, which may serve as a check on the interpretation in the case of Wh-question and Y/N-questions (where prosody is redundant vis-à-vis the lexico-syntactic information), or which is the sole information source the system has to go on (in the case of declarative questions).

## 5. REFERENCES


[1] G. Brown, K. Currie, J. Kenworthy, *Questions of intonation*, Croom Helm, London, 1980.

[2] S.J. Eady, W.E. Cooper, ‘Speech intonation and focus location in matched statements and questions’, *J. Acoust. Soc. Am.* 80, 1986, 402- 415.

[3] E. Hermann, *Probleme der Frage*, Nachrichten Akademie von Wissenschaft, Göttingen, 1942.

[4] D. Bolinger, ‘Nondeclaratives from an Intonational Standpoint’, in R. Schneider, K. Tuite, R. Chametzky, eds., *Papers from the parasession on nondeclaratives*, CLS, Chicago, 1982.

[5] G.A. Lindsey, *Intonation and interrogation: tonal structure and the expression of a pragmatic function in English and other languages*, Ph.D.-diss., UCLA, 1985.

[6] R. Geluijkens, ‘Questioning intonation’, *Antwerp Papers in Ling.*, 48, 1986.

[7] N. Thorsen, ‘A study of the perception of sentence intonation – evidence from Danish’, *J. Acoust. Soc. Am.*, 67, 1980, 1014-1030.

[8] J. Vaissière, ‘Language-independent prosodic features’, in A. Cutler, D. R. Ladd, eds., *Prosody: models and measurements*, Springer, Berlin, 1983.

[9] J. 't Hart, R. Collier, A. Cohen, *A perceptual analysis of intonation*, CUP, Cambridge, 1990.

[10] C. Gussenhoven, T. Rietveld, ‘A target-interpolation model for the intonation of Dutch’, *Proc. 2nd Int. Conf. Sp. Lang. Processing*, 2, 1992, 1235-1238.

[11] G.A. van Es, ‘Syntaxis en dialectstudie IV, intonatie en syntaxis [Syntax and the study of dialects IV, intonation and syntax]’, *Onze Taaltuin*, 1932, 122-128.

[12] J. Daan, ‘Dialect and pitch-pattern of the sentence’, *Proc. 3rd Int. Cong. Phon. Sc.*, 1938, 473-481.

[13] J. van Alphen, ‘De vraagzin [The question sentence]’, *De Nieuwe Taalgids*, 8, 1914, 88-95.

[14] J. Haan, V.J. van Heuven, J.J.A. Pacilly, R. van Bezooijen, ‘An anatomy of Dutch question intonation’, in H. de Hoop, M. den Dikken, eds, *Linguistics in the Netherlands 1997*, John Benjamins, Amsterdam (in press).

[15] G.N. Clements, ‘The status of register in intonation theory: comments on the papers by Ladd and by Inkelas and Leben’, in J. Kingston, M. Beckman, eds., *Papers in laboratory phonology I*, CUP, Cambridge, 1990, 58-71.

[16] D.J. Hermes, ‘Measurement of pitch by subharmonic summation’, *J. Acoust. Soc. Am.*, 83, 1988, 257-264.

[17] D.J. Hermes, J.C. van Gestel, ‘The frequency scale of speech intonation’. *J. Acoust. Soc. Am.*, 90, 1991, 97-102.

[18] C. Edelsky, ‘Question intonation and sex roles’, *Lang. in Soc*., 8, 1979, 15-32.

[19] P.A. Lachenbruch, *Discriminant analysis*, Hafner Press, New York, 1975.