On the role of the late rise and the early fall in the turn-taking system of Dutch

Johanneke Caspers

Universiteit Leiden

Abstract

The question posed in the present paper is whether subjects interpret a short utterance with a late non-prominent rise in pitch (LH%) as having a 'go on' function, prompting the current speaker to continue, whereas the same short utterance spoken with an accent-lending fall (H*L L%) is associated with finality, for example, with the answer to a yes-no question. A series of three perception experiments were run with natural data taken from Dutch Map Task dialogues. The results support the hypothesis that the LH% contour is associated with a 'go on' response, while the falling contour is associated with the answer to a question. Furthermore, LH% is preferred over H*L L% in contexts leading to backchannel responses, while there is no preference for either contour in question contexts. Finally, the LH% contour is acceptable in both context types, whereas the accent-lending fall is unacceptable in backchannel contexts.

1 Introduction

In everyday conversation there is generally a smooth and fast alternation of speaking turns, which can only be explained in terms of a highly complex system of interacting factors comprising syntax, semantics, pragmatics, prosody, visual cues, etc. My specific interest is in the function of one particular prosodic factor in the turn-taking process in Dutch, viz. speech melody.

In natural conversation so-called 'backchannels' (Yngve, 1970) are a common phenomenon. These are short optional utterances (for instance, *yes*, *hmhmm* or *okay*) produced by the current hearer to signal that (s)he is still engaged in the discourse, prompting the current speaker to go on. Communication is an interactional process, involving continuous feedback between interlocutors, and backchannels are important instances of responsive behavior. They signal that the information so far has been integrated into the common ground shared by speaker and listener (Clark & Brennan, 1991), and they also signal that the listener understands that the speaker has not finished yet. An utterance like *yes* can be used to indicate that the current listener has understood so far and that the speaker may continue with – for instance – giving directions. However, if the *yes* is a so-called conversational move, for example an answer to a yes-no question, it is not an optional utterance and therefore not a backchannel. It seems possible that the specific dialogue function of short utterances like *yes* is reflected in their suprasegmental characteristics.

In earlier investigations of a corpus of Dutch Map Task dialogues (task-oriented dialogues in which an 'instruction giver' has to explain to an 'instruction follower' how to draw a route on an unmarked copy of a map), backchannels were found to be often marked by a specific melodic configuration: a slight dip in pitch followed by a conspicuous rise, not lending overt prominence to the utterance, and therefore labeled as LH% (a label not present in the ToDI inventory, cf. Gussenhoven, Rietveld, Kerkhoff & Terken, 2003); the L stands for a low tone (not marking accent), and the H% for a high final boundary tone. The data revealed that the majority of short utterances functioning as encouraging background signals carry such a LH% contour, while lexically identical 'real' turns – generally answers to yes-no questions – were marked by a pitch accent in the majority of the cases. Figure 1 presents the percentage of LH% contours versus the percentage of pitch accents as marked by two expert labelers (for more details see Caspers, 2000, 2003a, under revision).

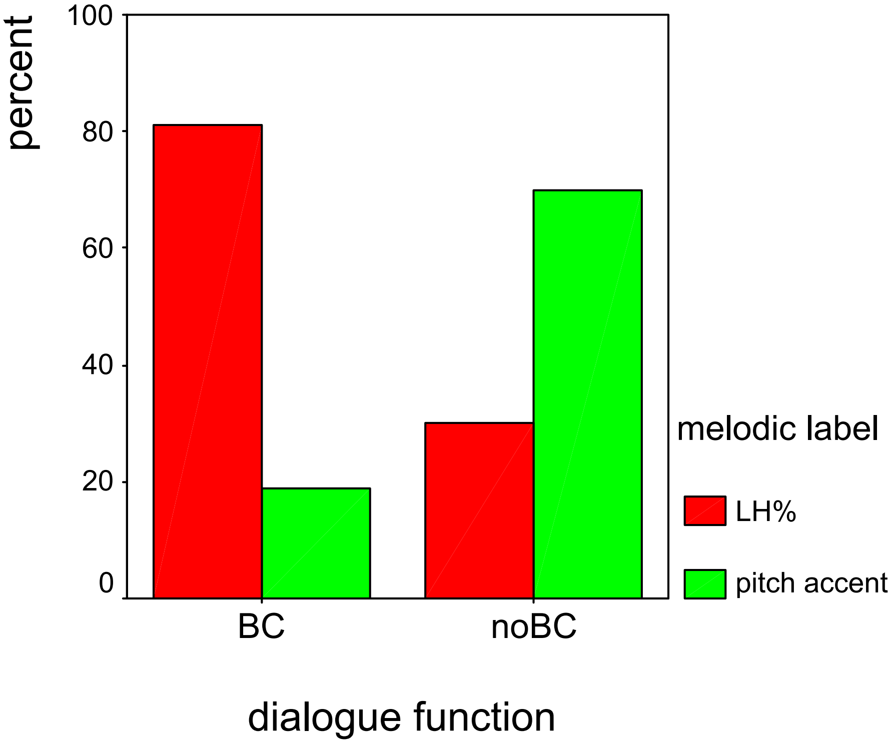

Figure 1. Percentage of LH% and pitch accent labels, broken down by dialogue function: BC (backchannel) versus noBC ('real' speaker turn).

This finding suggested that speech melody plays a role in signaling the dialogue function of short utterances like *yes* and *okay*, since there is a clear correspondence between backchannels and a non-prominent late rise and between 'real' speaker turns and a pitch accent. However, the LH% configuration does not seem to be an exclusive marker of backchannels, since it was found on approximately a third of the lexically identical 'real' turns as well. It could well be the case that LH% is essentially some sort of 'go on' signal, which suits backchannels in general, but may also fit certain 'real' speaker turns (for example, the answer to a yes-no question, which at the same time serves as an invitation to continue speaking; these kind of sequences are typical for Map Task dialogues, which essentially consist of one long instruction). The present perception experiment was designed to establish whether the LH% configuration is generally interpreted as a 'go on' signal in Dutch.

2 Approach

To be able to test the hypothesis that the LH% contour functions as a 'go on' signal in Dutch, it was contrasted with a contour that is supposedly not interpreted as such: the accent-lending fall (H*L L%), a contour typical for the positive answer to yes-no questions in the corpus materials, and associated with finality (Caspers, 1998, 1999, 2003b; Ladd, 1996; Rietveld & Gussenhoven, 1995).[1]

Figure 2 presents examples of a typical H*L L% and a typical LH% contour.

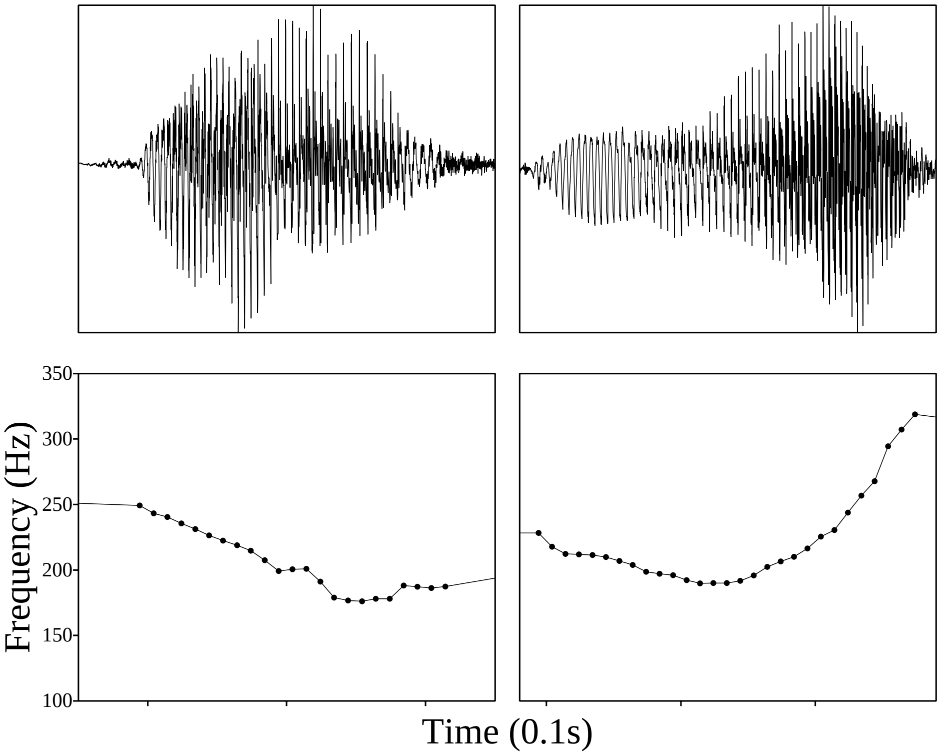

Figure 2. Examples of H*L L% (left) and LH% (right) contours on the word *ja* (yes); top: waveform, bottom: F_0-curve (in Hz).

Three sub-hypotheses were formulated:

1 Isolated utterances carrying LH% contours are associated with backchannels (since the main function of a backchannel is to signal to the speaker that the hearer is still there, which goes together naturally with 'go on'), in contrast with H*L L% contours, which will generally not be associated with a backchannel function.

2 In backchannel contexts there is a preference for LH% contours, in question contexts there is a preference for H*L L% contours.

3 LH% contours fit backchannel contexts as well as question contexts (because 'go on' is suitable as a positive reply to a yes/no question which is part of a larger instruction), while H*L L% contours will not fit backchannel contexts.

All combinations of the factors contour type (LH% vs. H*L L%) and dialogue function (BC vs. noBC) were available in the materials, albeit in different numbers. No manipulations of pitch were performed on the data, thereby preserving the naturalness of the stimuli. It also

[1] This fall is located early in the syllable, and is labeled 'A' in the Grammar of Dutch Intonation ('t Hart, Collier & Cohen, 1990). ToDI, the transcription system for Dutch intonation developed by Gussenhoven et al. (1999) uses the label H*L L% to refer to this type of contour, but H*L may also refer to a rising-falling pitch accent ('1&A' in the Grammar of Dutch Intonation).

means that, in addition to contour type, other prosodic and segmental information relevant to dialogue function may be present in the stimulus materials.

Only *ja*-utterances (yes) were used in the investigation, because they are the most frequently used lexical type of backchannel utterance (73% of the backchannels in the Map Task materials are *ja*'s), while they also occur most frequently as the (one-word) affirmative answer to a question (71%, Caspers, under revision).

For each hypothesis a separate experiment was conducted. To test hypothesis 1, isolated *ja*-utterances were presented to subjects, varying contour type (LH% versus H*L L%) as well as their original dialogue function (backchannel versus answer to a question). Subjects had to indicate whether they thought the *ja* was originally uttered as an optional background signal, prompting the current speaker to continue, or whether it was uttered as the positive answer to a yes-no question.

To test hypothesis 2, pairs of different *ja*-utterances were presented in a specific context, asking the subjects to select the utterance best fitting the given context. In the pairs of utterances either the contour type (LH% versus H*L L%), the original dialogue function (backchannel or answer), or both were contrasted; in each pair one of the two *ja*-utterances was originally produced in the presented context.

To test hypothesis 3, different combinations of a context and a *ja*-utterance were presented, asking subjects to rate the acceptability of each combination. As in the other parts of the experiment, contour type and dialogue function were systematically varied, leading to a combination of context and *original ja*-utterance in a quarter of the cases.

3 Method

3.1 Stimulus materials

All *ja*-utterances available in the Map Task materials (ca. 40 minutes of task-oriented Dutch dialogue, see Ladd, Schepman, Mennen & Lickley, 2001) were inspected for their usefulness in the present experiments. All cases of overlapping utterances were excluded, as well as all cases with immeasurable pitch contours.

The contexts were cut in such a way that they contained enough information for the subjects to determine the dialogue function of the immediately following utterance (i.e., to decide whether the *ja* was an optional backchannel or a non-optional answer to a yes-no question); their durations varied between 4.5 and 11 s.

3.2 Subjects

Twenty-four native speakers of Dutch participated in the experiment. They were paid a small fee. Fourteen were female, and their ages varied between 19 and 61.

3.3 Procedure

The interactive experiment was made accessible on the internet.[2] The contexts and stimuli to be judged were presented auditorily; subjects could press the relevant buttons as often as they found necessary.

Part A: subjects were presented with 24 different versions of *ja*. After (repeatedly) listening to each stimulus, they had to click a response button named either 'go on-signal' or 'answer

[2] The program for interactive stimulus presentation and response collection was written by Jos J.A. Pacilly of the Universiteit Leiden Phonetics Laboratory.

to a question'. The order of the stimuli and the spatial location of the two response buttons was blocked over subjects.

Part B: subjects were presented with 24 combinations of a context and two possible continuations, one of which they had to select as the best fitting. The order of the stimuli was reversed for half of the subjects.

Part C: subjects were presented with 16 combinations of a context and a *ja*-utterance, and judged the acceptability of each combination on a ten-point scale (in the Dutch educational system values 1 to 5 represent degrees of inadequacy, whereas values 6 to 10 represent degrees of adequacy; the boundary between acceptable and unacceptable is drawn at 5.5). The order of the stimuli was reversed for half of the subjects.

4 Results

4.1 Part A

In part A the subjects had to listen to a series of *ja*-utterances carrying either an LH% or an H*L L% contour, and indicate for each stimulus whether they thought it was uttered as a 'go on' signal or as the answer to a question, expecting an association between LH% contours and backchannels and between H*L L% contours and answers. The results are presented in table 1.

Table 1. Absolute (and relative) frequency of 'go on' and 'answer' responses for the LH% and H*L L% contours, broken down by original dialogue function (backchannel vs. answer to question).

		response		
contour type	original dialogue function	go on	answer	total
LH%	backchannel	117 (81%)	27 (19%)	144
	answer to question	125 (87%)	19 (13%)	144
	total	242 (84%)	46 (16%)	288
H*L L%	backchannel	52 (36%)	92 (64%)	144
	answer to question	26 (18%)	118 (82%)	144
	total	78 (26%)	210 (73%)	288
total		320 (56%)	256 (44%)	576

The table shows that in 84% of the cases a late-rising (LH%) contour is associated with a 'go on' function, while a falling contour (H*L L%) is associated with an answer in 73% of the cases (χ^2=189.11, *p*<.001), providing support for hypothesis 1. For the LH% contours there does not seem to be an additional influence of the original dialogue function of the stimulus: even when the stimulus functioned as the answer to a question in its original context, subjects associate it with a 'go on' function in 87% of the cases (χ^2=1.66, n.s.). However, for the accent-lending falls there does seem to be such an effect: when the stimulus originates from a *backchannel* context, the subjects associate it with an answer in 64% of the cases, but the association rises to 82% when the stimulus originally functioned as an answer to a question (χ^2=11.89, *p*<.001). This may mean that the association between an LH% contour and a 'go on' function is stronger than the association between H*L L% and the answer to a question. But it could also be the case that the *ja*-utterances carrying a falling contour contain more information referring to their original dialogue function than the utterances with a late rising contour.

4.2 Part B

In part B the subjects were presented with two different *ja*-utterances in a specific context. Their task was to select the one best fitting the given context, expecting the LH% contours to be preferred in backchannel contexts (part of a MapTask dialogue ending in a statement like *Then you take a right turn*) and the H*L L% contours to be preferred in answer contexts (a dialogue fragment ending in a question, e.g., *Do you have a stranded whale?*). Note that there is always a change of speaker between the end of the context and the *ja*-utterance.

In a third of the cases the two *ja*-stimuli from which the subjects had to choose did not differ in contour type, but only in their original dialogue function: backchannel or answer (and one of the two originally occurred in the presented context, as was the case for every stimulus pair). These cases were used to establish if subjects were able to hear which of the two stimuli is originally taken from the presented context. When the two *ja*'s differ in dialogue function (backchannel versus answer), but not in contour (*N*=192), there is a preference for the *ja* that was originally uttered in that context in 75% of the cases. This stimulus effect is larger than expected; it means that there is enough prosodic and/or segmental information available in the data for the subjects to connect the *ja* to its original context in three-quarters of the cases. At present it is not clear what kind of information this would be exactly.

The next question was: is there an *additional* effect of contour type? In table 2 the preference scores are presented for those cases where there was an opposition in contour type between the two alternative *ja*-utterances (*N*=384), making a distinction between utterances originating from the context presented and utterances originating from another context.

Table 2. Absolute (and relative) frequency of preference for LH% and H*L L% contours, for stimuli that were presented in their original context versus stimuli presented in a non-original context, broken down by context type.

	preference for contour			
	LH%		H*L L%	
context type	original	non-original	original	non-original
BC	88 (92%)	50 (52%)	46 (48%)	8 (8%)
noBC	68 (71%)	16 (17%)	80 (83%)	28 (29%)

Bearing in mind that there is a stimulus bias towards the original *ja*, an additional effect of contour type can be observed of approximately 20% for the LH% contours in backchannel contexts (a 92% preference when the LH% was produced in the presented context, i.e., 17% above the bias, and a 52% preference when the LH% was taken from another than the presented context, i.e., 27% above the bias). In the question (noBC) contexts the responses are roughly at the level of the stimulus bias (83% and 29%, which amounts to 8% and 4% above the stimulus bias respectively), which means that there is no clear additional influence of contour type in these cases. Presumably both contours fit these answers equally well and the original utterance – whose status as such is clear from other prosodic and/or segmental information – therefore wins. In the case of a backchannel there is indeed a preference for one of the two contour types, albeit a small one. This may mean that an LH% contour is acceptable as the answer to a question in the current materials.

4.3 Part C

In the third part of the test the subjects were presented with the same BC or noBC contexts, this time combined with just one *ja*-utterance, varying original dialogue function (BC or answer) and contour type (LH% and H*L L%). The subjects had to rate the acceptability of each combination of context and *ja* on a scale from 1 to 10. The LH% contour was expected to be acceptable in both context types, whereas the H*L L% contours were predicted to be acceptable in answer contexts only.

Table 3. Mean acceptability (and standard deviation) of contour type, broken down by context type.

	contour type presented		
context type	LH%	H*L L%	total
BC	6.1 (3.0)	4.8 (2.2)	5.4 (2.7)
noBC	5.7 (3.3)	6.3 (2.6)	6.0 (3.0)

The results presented in table 3 support the hypothesis that an LH% contour is acceptable in a backchannel context (a mean score of 6.1) and that an H*L L% contour is acceptable in a question context (6.3); furthermore, an LH% contour is – marginally – acceptable in a question context (5.7), whereas an H*L L% contour is clearly unacceptable in a backchannel context (4.8). Overall, the acceptability scores are rather low, and further inspection of the data (again) shows a large effect of the originality of the stimulus:

Table 4. Mean acceptability (and standard deviation) of contour type for stimuli that were presented in their original context versus stimuli presented in non-original context, broken down by context type.[3]

	contour type presented			
	LH%		H*L L%	
context type	original	non-original	original	non-original
BC	7.9 (1.9)	5.5 (3.1)	5.7 (2.4)	4.5 (2.0)
noBC	8.6 (1.4)	4.7 (3.1)	8.8 (1.2)	5.4 (2.4)
count	46	142	48	141

Table 4 reveals that stimuli presented in a non-original context are rated much lower than stimuli presented in their original contexts. An analysis of variance with fixed factors *contour type*, *originality of stimulus* and *context type* shows a small main effect of *context type* [*F*(1,375)=4.7, *p*<.05], a large effect of *originality* [*F*(1,375)=83.4, *p*<.001], interaction between *context type* and *contour* [*F*(1,372)=13.1, *p*<.001] and between *context type* and *originality* [*F*(1,372)=9.2, *p*<.005].

The main effect of context type – *ja*'s presented in question contexts are on average perceived as more acceptable (a mean score of 6.0) than *ja*'s in backchannel contexts (a mean score of 5.4) – could well be caused by the fact that some of the backchannel contexts are clearly unfinished (in content, but sometimes prosodically as well), while the question contexts are neatly finished off by a positive reply.

The main effect of originality was not expected to be this large: original *ja*'s get a mean acceptability score of 7.8 (*sd* 2.2), while non-original *ja*'s have an average acceptability score of 5.0 (*sd* 2.7).

[3] Because of a mistake by one of the subjects, seven cases are missing from the dataset.

The interaction between *context type* and *contour type* was predicted, but the main effect and interaction regarding the originality of the stimuli were not. However, there is no three-way interaction between *context type*, *contour type* and *originality*, which indicates that the predicted effect is present in the data, irrespective of the influence of the originality of the stimulus.

Support for hypothesis 3 can only be found for the stimuli presented in their original contexts: stimuli with LH% contours are amply acceptable in backchannel contexts as well as question contexts (mean scores of 7.9 and 8.6, respectively), whereas stimuli with a falling contour are fully acceptable *only* in a question context (a mean score of 8.8); when appearing in a backchannel context, they are judged as – comparatively – unacceptable, despite the fact that the stimulus was presented in its original context (a mean score of 5.7). For the stimuli presented in non-original contexts there is a trend toward higher acceptability for LH% contours in backchannel contexts and for H*L L% contours in question contexts, but the overall acceptability of these stimuli is below 6. As in parts A and B, this means that the stimuli must contain information that reveals that they were taken from another context than they were presented in.

5 Discussion and conclusion

Summarizing the results, subjects clearly associate a late-rising contour (LH%) with a 'go on' function and an early accent-lending fall (H*L L%) with an answer to a question, the latter association being a little weaker. Furthermore, subjects prefer an LH% contour over an H*L L% contour in a context leading to a backchannel, which is compatible with the proposed function of contour LH% as signalling 'go on', while there is no clear preference for either contour type in a question context. Finally, an LH% contour is acceptable in both context types, whereas the accent-lending fall is unacceptable in a backchannel context (even if it was originally produced there), supposedly because the fall puts the *ja* itself in focus, and because the fall is associated with finality, which does not suit a real backchannel very well.

The results showed an unexpectedly large influence of prosodic and/or segmental characteristics other than contour type: in the turn-changing contexts there was no clear preference for a specific contour type, but there was a clear preference for the stimulus originally uttered in that specific context, and only the original stimuli were judged to be generally acceptable. This means that subjects were able to determine whether or not a stimulus was presented in its original context in a majority of the cases, probably on the basis of varying combinations of prosodic characteristics of the stimulus itself (voice quality, loudness contour, pitch range, duration, segmental characteristics, etc.), as well as information contained in the transition between the end of the presented context and the following stimulus (despite the fact that there was always an intervening pause).

However, the predicted association between LH% and 'go on' is clearly visible in the data, which may well explain why the LH% contour, which typically appears on backchannels, may also appear on certain non-optional 'real' turns. In contrast, the accent-lending fall does not seem to be a very appropriate contour for backchannels, supposedly because this contour is associated with prominence and finality.

The proposed functions of the two contours investigated appear to be opposites at first sight, but they certainly do not exclude one another. In light of this fact the presented results seem quite clear.

References

Caspers, J. (1998). Experiments on the meaning of two pitch accent types: the ‘pointed hat’ versus the accent-lending fall in Dutch. In *Proceedings of the International Conference on Spoken Language Processing, Sydney* (pp. 1291-1294).

Caspers, J. (1999). The early versus the late accent-lending fall in Dutch: Phonetic variation or phonological difference. In *Proceedings of the International Congress of Phonetics Science*s*, San Francisco* (pp. 945-948).

Caspers, J. (2000). Melodic characteristics of backchannels in Dutch Map Task dialogues. In *Proceedings of the International Conference on Spoken Language Processing, Beijing* (Vol. II, pp. 611-614).

Caspers, J. (2003a). Local speech melody as a limiting factor in the turn-taking system in Dutch. *Journal of Phonetics*, *31*, 251-276.

Caspers, J. (2003b). Phonetic variation or phonological difference? The case of the early versus the late accent-lending fall in Dutch. In: J. van de Weijer, V.J. van Heuven & H. van der Hulst (Eds.), *The Phonological Spectrum. Volume II: Suprasegmental structure* (pp. 201-223). Amsterdam/Philadelphia: John Benjamins.

Caspers, J. (under revision). Melodic characteristics of backchannels in Dutch task-oriented dialogues. *Speech Communication*.

Clark, H.H., & Brennan, S.E. (1991). Grounding in communication. In: L.B. Resnick, J.M. Levine, & S.D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington DC: American Psychological Association.

Gussenhoven, C., Rietveld, T., Kerkhoff, J., & Terken, J. (2003). *ToDI, Transcription of Dutch Intonation* (second edition). Available: http://todi.let.kun.nl/ToDI/home.htm.

Hart, J. ‘t, Collier, R., & Cohen, A. (1990). *A perceptual study of intonation*. Cambridge: Cambridge University Press.

Ladd, D.R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

Ladd, D.R., Schepman, A., Mennen, I., & Lickley, R.J. (2001). Final report on research activities and results for ESRC project No. R000-23-7447 ‘Alignment of Fundamental Frequency Targets in English and Dutch’. Available: www.ling.ed.ac.uk/eprints/.

Rietveld, T., & Gussenhoven, C. (1995). Aligning pitch targets in speech synthesis: effects of syllable structure. *Journal of Phonetics, 23*, 375-385.

Yngve, V. (1970). On getting a word in edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society.