ON THE FUNCTION OF THE LATE RISE AND THE EARLY FALL IN DUTCH DIALOGUE: A PERCEPTION EXPERIMENT

Johanneke Caspers

Phonetics Laboratory, Universiteit Leiden Centre for Linguistics
Cleveringaplaats 1, 2311 BD Leiden, The Netherlands
j.caspers@let.leidenuniv.nl

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% or 'A') 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, and 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 A 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: speech melody.

In natural conversation so-called 'backchannels' [1] are a common phenomenon: short optional utterances 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 interactive process, involving continuous feedback between interlocutors, and backchannels are an important instance of responsive behavior. They signal that the information so far has been integrated into the common ground shared by speaker and listener [2], 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 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 Dutch task-oriented dialogue the majority of backchannels were found to be 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% [3,4,5]. The data revealed that short utterances functioning as stimulating background signals carry a LH% contour in 69% of the cases, while lexically identical ‘real’ turns – mostly answers to yes-no questions – were marked by a pitch accent in 61% of the cases (typically H*L L%, see figure 1 for an example of an H*L L% and an LH% contour). This finding suggested that speech melody plays a role in signaling the dialogue function of short utterances like ‘yes’ and ‘okay’, which is in line with earlier reported results for Dutch disconfirmations [6].

![Figure 1: Examples of H*L L% (left) and LH% (right) contours on the word “ja” ('yes'); above: waveform, below: F0 curve in Hz.](image)

However, the LH% configuration does not seem to be an exclusive marker of backchannels, since it was found on approximately a quarter 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 fit certain ‘real’ speaker turns as well (for example, the answer to a yes-no question, which at the same time serves as an invitation to continue speaking). 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, a contour typical for the positive answer to yes-no questions, and associated with finality [7,8,9,10,11]. This fall is located early in the syllable, and is labeled ‘A’ in the Grammar of Dutch Intonation [12]. ToDI, the transcription system for
Dutch intonation developed by Gussenhoven, Rietveld and Terken [13] 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), and therefore the label ‘A’ is used in the present paper.

Three sub-hypotheses were formulated:

1. LH% contours are associated with backchannels, in contrast with A 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 A.

3. LH% contours fit backchannel contexts as well as question contexts, while A contours will not fit backchannel contexts.

The materials used for testing the hypotheses were taken from the Map Tasks used in the corpus investigation [3,4,5]. This means that no manipulations of pitch were performed on the data, thereby preserving the naturalness of the stimuli. It also means that, in addition to contour type, other prosodic information relevant to dialogue function may be present in the stimulus materials (duration, loudness, etc.). However, no systematic effects of dialogue function (backchannel versus answer to question) on prosodic variables besides intonation contour type were found in the investigated materials [5].

Only ‘ja’-utterances (‘yes’) were used in the present investigation, because they are the most frequently used form of backchannel utterance (73%), while they also occur most frequently as the – one-word – affirmative answer to a question (71%) [5].

For each hypothesis a separate experiment was conducted. To test hypothesis 1, isolated ‘ja’-utterances were presented to subjects, varying contour type (LH% versus A) as well as their original dialogue function (backchannel versus answer to a question). Subjects had to indicate whether they thought the ‘ja’ was 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 A), the original dialogue function (backchannel or answer), or both were contrasted; in each pair one of the two ‘ja’-utterances was the original one.

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 only a quarter of the cases.

2.1. METHOD

2.1.1. Stimulus Materials

All ‘ja’-utterances available in the Map Task materials (ca. 40 minutes of task-oriented dialogue) were inspected for their usefulness in the present experiments. All cases of overlap were excluded, as well as all cases with immeasurable pitch contours. The combination of a falling pitch accent (A) and a backchannel function occurred in only 6 utilizable cases (as opposed to e.g. 103 utilizable cases where a backchannel was marked by an LH% contour), which limited the set of stimuli to 24.

The contexts were cut in such a way that they contained enough information to determine the dialogue function of the immediately following utterance, ‘ja’ (i.e., functioning as an optional backchannel or as a nonoptional answer to a yes-no question); their durations varied between ca. 4.5 and 11 s.

2.1.2. Subjects

Potential subjects were approached via email. 24 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.

2.1.3. Procedure

The interactive experiment was presented on the internet (HTTP://FONETIEK-6.LEIDENUNIV.NL/ASPIERS/LE6-INTRO.HTML, programmed by Ing Jos J.A. Pacilly of the Universiteit Leiden Phonetics Laboratory). 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 to a question’. The order of the stimuli and the order 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. In addition, subjects had to indicate whether they were sure or unsure about their preference. 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 asked to judge 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, and the boundary between acceptable and unacceptable is drawn at 5.5). The order of the stimuli was reversed for half of the subjects.

3. RESULTS

3.1. PART a

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

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 (A) is associated with an answer in 73% of the cases ($\chi^2=189.11, p<0.001$), supporting 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 backchannel function in 87% of the cases ($\chi^2=1.66$, ins.). 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 only 64% of the cases, while the association rises to 82% when the stimulus originally functioned as an answer (χ² = 11.89, p<.001). This indicates that the association between LH% and a ‘go on’ function is stronger than the association between A and the answer to a question.

<table>
<thead>
<tr>
<th>contour type</th>
<th>dialogue function</th>
<th>response</th>
<th>go on</th>
<th>A</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>LH%</td>
<td>backchannel</td>
<td>turn change</td>
<td>117 (81%)</td>
<td>27 (19%)</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>125 (87%)</td>
<td>19 (13%)</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>total</td>
<td></td>
<td>242 (84%)</td>
<td>46 (16%)</td>
<td>288</td>
</tr>
<tr>
<td>A</td>
<td>backchannel</td>
<td>turn change</td>
<td>52 (36%)</td>
<td>92 (64%)</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>26 (18%)</td>
<td>118 (82%)</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>total</td>
<td></td>
<td>78 (26%)</td>
<td>210 (73%)</td>
<td>288</td>
</tr>
</tbody>
</table>

Table 1: Absolute (and relative) frequency of ‘go on’ and ‘answer’ responses for the LH% and A contours, broken down by original dialogue function (backchannel vs. change of turn).

3.2. 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 and the A contours to be preferred in answer contexts.

<table>
<thead>
<tr>
<th>context type</th>
<th>preference</th>
<th>LH%</th>
<th>A</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>backchannel</td>
<td></td>
<td>138 (72%)</td>
<td>54 (28%)</td>
<td>192</td>
</tr>
<tr>
<td>turn change</td>
<td></td>
<td>84 (44%)</td>
<td>108 (56%)</td>
<td>192</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>222 (58%)</td>
<td>162 (42%)</td>
<td>384</td>
</tr>
</tbody>
</table>

Table 2: Absolute (and relative) frequency of preference for LH% and A contours, broken down by context type (for those cases where there is an opposition between contour types).

The table shows a preference for the LH% contour in 72% of the backchannel contexts, while the preference for the falling contour in answer contexts is only 56% (χ² = 31.14, p<.001).

Table 3 presents the original dialogue function of the preferred stimuli in the two context types:

<table>
<thead>
<tr>
<th>context type</th>
<th>preference</th>
<th>backchannel</th>
<th>no-backchannel</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>backchannel</td>
<td></td>
<td>126 (66%)</td>
<td>66 (34%)</td>
<td>192</td>
</tr>
<tr>
<td>turn change</td>
<td></td>
<td>23 (12%)</td>
<td>169 (88%)</td>
<td>192</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>149 (39%)</td>
<td>235 (61%)</td>
<td>384</td>
</tr>
</tbody>
</table>

Table 3: Absolute (and relative) frequency of preference for backchannel versus no-backchannel, broken down by context type (for those cases where there is an opposition in original dialogue function).

When there is an opposition between the two stimuli in original dialogue function (note that the cases contrasting melody as well as function are included in tables 2 and 3), the data show a moderate preference for backchannel stimuli in backchannel contexts (66%) and a strong preference for original answers in question contexts (88%, χ² = 116.35, p<.001). It seems that subjects have a clear preference for an LH% contour in a context where a backchannel is appropriate, whereas they have no preference for either contour type in a turn-changing context; yet they are able to hear which of two stimuli originates from the presented context, and that is the one they prefer.

Subjects are sure of their preference more often than unsure (72% vs. 28%), and there is no effect of context type (χ² = 3.80, ins.). There is, however, a small effect of the type of opposition between the two stimuli they have to choose from:

<table>
<thead>
<tr>
<th>opposition type</th>
<th>preference</th>
<th>unsure</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>melody</td>
<td>142 (74%)</td>
<td>50 (26%)</td>
<td>192</td>
</tr>
<tr>
<td>function</td>
<td>125 (65%)</td>
<td>67 (35%)</td>
<td>192</td>
</tr>
<tr>
<td>both</td>
<td>148 (77%)</td>
<td>44 (23%)</td>
<td>192</td>
</tr>
<tr>
<td>total</td>
<td>415 (72%)</td>
<td>161 (28%)</td>
<td>576</td>
</tr>
</tbody>
</table>

Table 4: Absolute (and relative) frequency of sure and unsure preference responses, broken down by opposition type.

Subjects are most unsure when they have to choose between stimuli differing in original function only (χ² = 7.36, p<.05).

As predicted, an LH% contour is preferred in backchannel contexts. However, there is no clear preference for a stimulus carrying an accent-lending fall in a turn-changing context, which may mean that an LH% contour is perfectly acceptable as the answer to a question in the current materials.

3.3. PART c

In the third part of the test the subjects were presented with a context plus a 'ja'-utterance and they had to rate the acceptability of the combination, expecting the LH% contours to be acceptable in both context types, whereas the A contours were predicted to be acceptable in answer contexts only. Because of a mistake of one of the subjects, seven cases are missing from the dataset.

<table>
<thead>
<tr>
<th>context type</th>
<th>contour type presented</th>
<th>LH%</th>
<th>A</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>backchannel</td>
<td></td>
<td>6.1 (3.0)</td>
<td>4.8 (2.2)</td>
<td>5.4 (2.7)</td>
</tr>
<tr>
<td>turn change</td>
<td></td>
<td>5.7 (3.3)</td>
<td>6.3 (2.6)</td>
<td>6.0 (3.0)</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Mean acceptability (and standard deviation) of contour type, broken down by context type.

The results presented in table 5 support the hypothesis that an LH% contour is acceptable in a backchannel context (a mean score of 6.1) and that an A contour is acceptable in a question context (6.3); furthermore, an LH% contour is – marginally – acceptable in a question context (5.7), whereas an A contour is clearly unacceptable in a backchannel context (4.5). Overall the acceptability scores are rather low, and further inspection of the data shows a large effect of the originality of the stimuli:

<table>
<thead>
<tr>
<th>context type</th>
<th>contour type presented</th>
<th>LH% original</th>
<th>non-original</th>
<th>A original</th>
<th>non-original</th>
</tr>
</thead>
<tbody>
<tr>
<td>backch.</td>
<td></td>
<td>7.9 (1.9)</td>
<td>5.5 (3.1)</td>
<td>5.7 (2.4)</td>
<td>4.5 (2.0)</td>
</tr>
<tr>
<td>turn ch.</td>
<td></td>
<td>8.6 (1.4)</td>
<td>4.7 (3.1)</td>
<td>8.8 (1.2)</td>
<td>5.4 (2.4)</td>
</tr>
<tr>
<td>count</td>
<td></td>
<td>46</td>
<td>142</td>
<td>48</td>
<td>141</td>
</tr>
</tbody>
</table>

Table 6: 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.
Table 6 reveals that stimuli presented in a non-original context are rated much lower than stimuli presented in 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 type ($F_{(1,375)} = 13.1, p < .001$) and between context type and originality ($F_{(1,375)} = 9.2, p < .005$). The interaction between context type and contour type was predicted, but the main and interaction effects 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 each 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 A contours in question contexts, but the overall acceptability of these stimuli is below 6. As in part b, this means that the stimuli must contain information that reveals that they were taken from another context than they are presented in, and this information has to be prosodic in nature.

4. DISCUSSION AND CONCLUSION
Summarizing the results, subjects clearly associate a late-rising contour (LH%) with a ‘go on’ function and an accent-lending fall (A) with an answer to a question, the latter association being a little weaker. Furthermore, subjects prefer an LH% contour over an A contour in a context leading to a backchannel, 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). This means that the LH% contour is indeed interpreted as a signal that the other speaker should continue, as opposed to the falling contour.

The results of parts b and c showed an unexpectedly large influence of prosodic 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 original stimulus, and only the original stimuli were judged to be generally acceptable. This means that subjects were able to determine whether a stimulus was presented in its original context, probably on the basis of varying combinations of prosodic characteristics of the stimulus itself (voice quality, loudness contour, pitch range, duration, etc.), as well as information contained in the connection 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, that 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 finality.

5. REFERENCES

ACKNOWLEDGMENTS
This research was funded by the Netherlands Organization for Scientific Research (NWO), under project #355-75-002.