MELODIC CHARACTERISTICS OF BACKCHANNELS IN DUTCH MAP TASK DIALOGUES

Johanneke Caspers

Phonetics Laboratory/Holland Institute of Generative Linguistics
PO Box 9515, 2300 RA Leiden, The Netherlands; CASPERS@RULLET.LEIDENUNIV.NL

ABSTRACT

In natural conversation backchannels (short optional utterances like 'uh-huh' or 'yes') are used to indicate to the current speaker that the current listener understands so far and that the speaker may continue. The question posed in the present paper is whether backchannels distinguish themselves melodically from lexically identical utterances that have a different function (e.g., the answer to a question). In a corpus of Dutch Map Task dialogues the melodic configurations realized on all backchannels and all lexically identical non-backchannels were transcribed using ToDI [1]. Comparison of the two groups of data reveal a clear tendency for backchannels to be marked by a non-prominence lending drop in pitch followed by a high boundary tone (69%), whereas the non-backchannels generally carry a pitch accent (61%).

1. INTRODUCTION

In natural conversation so-called 'backchannels' (the term was introduced by Yngve [2], cf. [3]) are a common phenomenon: short optional utterances produced by the hearer to signal that s/he is still engaged in the dialogue, 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. As the name indicates, backchannels are not viewed as speaker turns, but as sounds occurring *during* the turn of another speaker (and they are normally left out when a conversation is reproduced). For instance, 'yes' can be used to indicate that the current listener has understood so far and that the speaker may continue. However, if '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 structure.

Recent research within the framework of human-machine communication has taken a growing interest in the function of prosody in human-human conversation, with the aim of improving the performance of existing spoken dialogue systems. For example, investigations of the prosodic characteristics of echoic responses in Japanese dialogues [4,5] show that the function of these repeated parts of speech, such as acknowledgment or repair-request, tends to be reflected in their prosodic characteristics. Likewise, investigations of Dutch disconfirmations revealed close connections between the function of the word "nee" ('no', which serves as a 'go on' or a 'go back' signal in the investigated corpus) and it's prosodic characteristics [6]. These findings may induce the assumption that backchannels – by definition affirmative in nature, i.e., 'go on' signals – have specific prosodic characteristics, such as a short duration, a short preceding pause, and no marked pitch accent.

The question posed in the present paper is whether backchannels distinguish themselves melodically from lexically identical 'real' turns. To answer this question, backchannels and lexically identical non-backchannels were marked in a corpus of Dutch task-oriented dialogues, and the melodic characteristics of both utterance types were compared.

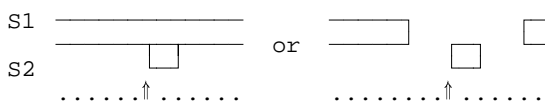## 2. METHOD

2.1. Materials

Use was made of a corpus of Dutch guided spontaneous conversations, so-called Map Task dialogues (cf. [7]). In these task-oriented dialogues, maps provide a handle on an essentially spontaneous conversation. Two roles can be distinguished: an 'instruction giver' and an 'instruction follower'. The former participant has a map with a route drawn on it and s/he has to explain to the instruction follower which route to draw on his or her unmarked copy of the map. The participants cannot see each other's maps. Both maps have a number of reference points on it (e.g., 'old pond', 'new pond', 'green meadow') and by introducing small differences between these reference points it is possible to complicate the dialogue to some extent. The materials used for the present investigation were collected by Bob Ladd and Astrid Schepman.

2.2. Analysis

Inter Pausal Units. The materials were divided into so-called *Inter Pausal Units*. An IPU is defined as "a stretch of a single speaker's speech bounded by pauses longer than 100 ms" ([8], p. 299). This means that boundaries were drawn in all positions where a pause longer than 100 ms appeared in the signal, and in positions where a change of speaker occurred. IPUs can be determined objectively and the boundaries between these units can be labeled as instances of either turn-holding or turn-changing.

Backchannels. Every IPU was labeled in turn transitional terms (cf. [9] for a detailed description of the labeled data), using the definition for backchannels as formulated by Koiso et al. [8]). Backchannels occur either during the speech produced by the current speaker, or in a pause made by the

current speaker, see below (the boxes following S1 depict the stretches of speech uttered by the current speaker, S2 indicates the speech by the listener, the dotted line marks the time course, and the arrow indicates the relevant IPU boundary):

```
S1 ─────────          ┌───┐      ┌─┐
     ─────┌─┐  or  ───┘     └─┐  ┌─┘ └─┐  ┌──
S2       └─┘                └─┘          └──
   ......⇑......      .......⇑......
```

In a minority of the cases it was difficult to decide whether a specific IPU was a backchannel or not, because the optionality of the utterance was hard to assess. As was already mentioned by Schegloff in his study of backchannels ([10], p. 85), speakers sometimes actually wait for their listener to produce a continuer like 'uh-huh', which means that some backchannels are more or less obligatory in nature. Note that the labeling of the backchannels was done on the basis of an orthographic transcription of the materials, without using any acoustic cues; only when the orthography gave insufficient information as to whether or not the preceding IPU should be interpreted as a question (and, hence, the IPU under investigation as the answer to a question), the melody of the preceding IPU was taken into account.

Below a part of a Map Task is represented, containing five short utterances by the instruction follower (S). IPUs number 2, 8 and 10 were labeled as backchannels, whereas 4 and 6 were labeled as 'real' turns, since they form the answer to a yes-no question produced by the instruction giver (X).

1. (X) Dus voordat het naar links gaat buigen stop jij met de lijn
        'So before it bends to the left you stop the line'

2. (S) Ja
        'Yes'

3. (X) En dan... heb jij 't beeld van oorlogsheld
        'And then... do you have the statue of war hero'

4. (S) Huhum
        'uh-huh'

5. (X) Rechts
        'To the right'

6. (S) Ja
        'Yes'

7. (X) Daar ga je recht naar toe vanaf daar
        'There you go in a straight line from there'

8. (S) Oké
        'Okay'

9. (X) En daar ga je onderlangs omheen
        'And there you go underneath around'

10. (S) Ja
        'Yes'

Intonation Transcription. As a tool for labeling the melodic phenomena, the recently developed ToDI system ('Transcription of Dutch Intonation', [1]) was used. The last pitch accent – when present – before every IPU boundary was transcribed, as well as the tone sequence following this accent up to the boundary. The intonation was labeled on the basis of the auditory impression of the pitch curve only. Before every IPU boundary a boundary tone was transcribed, so that intonation domain boundaries were determined by pauses or speaker changes actually occurring in the material, and not by the syntactic structure of the utterance. Note that, as a result, the boundary tones marked in the current analysis do not necessarily correspond to the boundary tones as defined by ToDI.

The intonation of overlapping stretches of speech could not always be labeled properly, which led to a small amount of 'intranscribable' data ("?"). Another issue was the transcription of a melodic configuration that seemed typical for backchannels: a clear drop in pitch immediately followed by a rise, without the overt suggestion of prominence, which means that these configurations could not be interpreted as pitch accents. ToDI does not offer a suitable label (neither does the Grammar of Dutch Intonation, cf. [11]) and therefore I decided to give these specific configurations the label LH% (i.e., a low tone followed by a high boundary tone). Figure 1 presents examples of a default falling pitch accent (H*L) and of an LH% contour on the word "ja" ('yes'), uttered by the same female instruction follower.
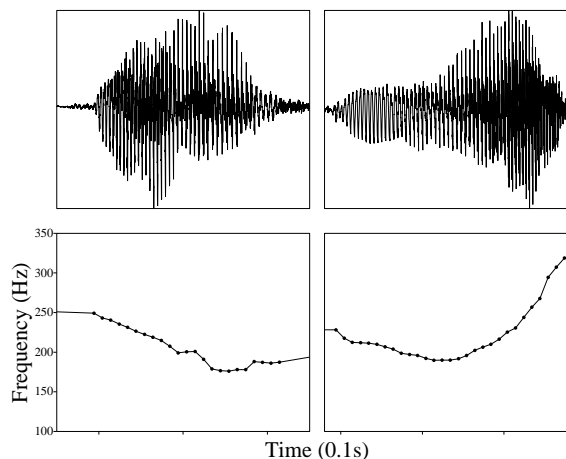

Figure 1: Example of H*L L% (left) and LH% (right) contours on the word "ja" ('yes'); above: waveform, below: F_0 curve (in Hz).

2.3. Expectations

A tendency was expected for backchannels to carry LH% contours, whereas the lexically identical non-backchannels were expected to be marked by clear pitch accents. Eight complete Map Task dialogues were analyzed, amounting to o-

lexical shape	melodic shape			total
	pitch accent	LH%	?	
"ja" ('yes')	31 (22%)	95 (69%)	12 (9%)	138 (73%)
"oké" ('okay')	7 (29%)	15 (63%)	2 (8%)	24 (13%)
other	3 (11%)	20 (74%)	4 (15%)	27 (14%)
total	41 (22%)	130 (69%)	18 (9%)	189 (100%)

Table 1: Absolute (and relative) frequency of occurrence of the different melodic shapes of the backchannels encountered in the material, broken down by lexical shape.

ver 40 minutes of speech. These materials contained 1552 IPU boundaries, among which were 189 backchannels. There were 153 instances of lexically identical 'real' speaker turns.

3. RESULTS

Table 1 presents the lexical shape and the general melodic characteristics of the 189 backchannels encountered in the material. In 73% of the cases the listener uttered a simple "ja" ('yes'); a further 13% of the cases existed of the utterance "oké" ('okay') and in the remainder of the cases "hmhmm" ('uh-huh'), "oh" ('oh') or "goed" ('good') were used. The backchannels in the investigated material carry a low tone followed by a conspicuous final rise in almost 70% of the cases (LH%), while the remaining cases could either not be transcribed ("?", 9%) or were marked by a pitch accent (H*, H*L, L* or L*H) followed by a low (L%), high (H%) or level (%) boundary tone (22%). For lack of space the different types of pitch accents and boundary tones were collapsed into one category.

Table 2 contains the lexical shape and melodic characteristics of the non-backchannels that were found in the material. The table reveals that the 153 turns resembling backchannels have a similar lexical distribution (71% 'yes', 23% 'okay' and 6% 'oh', 'uh-huh' or 'good'). Their melodic characteristics, however, differ clearly: the non-backchannels are marked by a pitch accent in 61% of the cases and carry a LH% configuration in only 27% of the cases; the remaining 12% could not be transcribed.

An ANOVA was performed on the percentage of backchannels versus non-backchannels with melodic type (the four main pitch accent types – H*, H*L, L* and L*H – plus the categories LH% and "?") and lexical shape ("ja", "oké" and "other") as factors. There were main effects of melodic type ($F_{(5,336)}$= 15.51, p<.001) and lexical shape ($F_{(2,339)}$= 3.51, p<.05) on the distribution of backchannels versus non-backchannels, but no interaction ($F_{(10,324)}$= 1.72, ins.). A posthoc analysis shows that LH% differs significantly from all other melodic types, except from the configuration L*H (a pitch accent type that occurs only 6 times in the data and closely resembles LH%); there were no further differences.

As expected, the backchannels present in the current materials are more often marked by LH% than the lexically identical IPUs that constitute an actual speaker turn, and, vice versa, the 'real' speaker turns are more often marked by a pitch accent.

4. OTHER PROSODIC CUES?

The melodic characteristic of a short utterance like 'yes' is probably not the only prosodic cue to its function. As was reported by Krahmer et al. [6], Dutch 'go on' signals are generally shorter than 'go back' signals, and they are preceded by a shorter pause.

Investigation of the current data, however, shows no main effect of the BC-NOBC opposition on the duration of the relevant IPUs ($F_{(1,340)}$= 2.53, ins.) and no interactions with other independent variables. It thus seems that there is no inherent difference in duration between backchannels and non-backchannels, which probably means that the 'real' turns in the current materials cannot be compared to the 'go back' signals investigated by [6].

Further investigation indicates that the BC-NOBC opposition does influence the duration of the preceding pause: it is indeed shorter when it precedes a backchannel than when it precedes a turn ($F_{(1,340)}$= 4.23, p<.05). However, this effect is reversed when the pause precedes "oké" (there is a main effect of lexical shape, $F_{(2,339)}$= 3.94, p<.05 and a significant interaction, $F_{(2,336)}$= 5.64, p<.01). The reason for this reversal may be that 'okay' signals a more hesitant type of affirmation than 'yes' or 'uh-huh'.

Summarizing, the investigated materials do not present simple durational cues to the function of short utterances like 'yes'.

5. CONCLUSION AND DISCUSSION

The data suggest that speech melody plays a role in signaling the dialogue function of short utterances like 'yes' and 'okay'. The data show a clear tendency for backchannels to carry a LH% contour (69%) and for segmentally identical 'real' turns to be marked by a pitch accent (61%). The melodic shape thus

lexical shape	melodic shape			total	
	pitch accent	LH%	?		
"ja" ('yes')	62 (57%)	34 (32%)	12 (11%)	108	(71%)
"oké" ('okay')	26 (72%)	6 (17%)	4 (11%)	36	(23%)
other	5 (56%)	1 (11%)	3 (33%)	9	(6%)
total	93 (61%)	41 (27%)	19 (12%)	153	(100%)

Table 2: Absolute (and relative) frequency of occurrence of the melodic shape of non-backchannels, broken down by lexical shape.

seems to support information that can be derived from other sources, such as syntactic and pragmatic structure.

However, it should be noted that the materials investigated are quite specific in nature: Map Task dialogues consist mainly of instructions, with occasional questions and checks uttered by the instruction follower. It could be the case that other types of dialogue elicit backchannels with different melodic characteristics. Further investigation of diverse types of spontaneous conversation will have to reveal whether or not backchannels in general typically carry a non-prominence lending low tone followed by a high boundary tone.

Furthermore, the LH% configuration does not seem to be an exclusive marker of backchannels, since it was found on approximately a quarter of the '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 normal speaker turns as well (for example, the answer to a yes-no question, which at the same time serves as a continuation signal, cf. line 6 in the example given in subsection 2.2.). Further perception experiments are needed to establish whether the LH% configuration is generally interpreted as a 'go on' signal in Dutch.

6. ACKNOWLEDGMENTS

7. REFERENCES

1. Gussenhoven, C., Rietveld, T. and Terken, J., "ToDI, Transcription of Dutch Intonation", URL http://lands.let.kun.nl/todi/, 1999.

2. Yngve, V., "On getting a word in edgewise", *Papers from the Sixth Regional Meeting, Chicago Linguistic Society*, Chicago, 567-577, 1970.

3. Duncan, S. and Fiske, D.W., *Face-to-face interaction: Research, methods, and theory*, Lawrence Erlbaum Associates, Hillsdale, N.J., 1977.

4. Swerts, M., Koiso, H., Shimojima, A., and Katagiri, Y., "On different functions of repetetive utterances", *Proceedings of the 5th International Conference on Spoken Language Processing*, Sydney, 483-486, 1998.

5. Shimojima, A., Katagiri, Y., Koiso, H., and Swerts, M., "An experimental study on the informational and grounding features of Japanese echoic responses", *Proceedings of the ESCA Workshop on Dialogue and Prosody*, Eindhoven, 187-192, 1999.

6. Krahmer, E., Swerts, M., Theune, M., and Weegels, M., "Prosodic correlates of disconfirmations", *Proceedings of the ESCA Workshop on Dialogue and Prosody*, Eindhoven, 169-174, 1999.

7. Anderson, A.H., Bader, M., Gurman Bard, E., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., and Thompson, H.S., "The HCRC Map Task Corpus", *Language and Speech* 34, 351-366, 1991.

8. Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., and Den, Y., "An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese Map Task dialogs", *Language and Speech* 41, 295-321, 1998.

9. Caspers, J., "Looking for melodic turn-holding configurations in Dutch", in H. de Hoop and T. van der Wouden (eds), *Linguistics in the Netherlands 2000*, John Benjamins, Amsterdam, to appear.

10. Schegloff, E.A., "Discourse as an interactional achievement: some uses of 'uh huh' and other things that come between sentences", in D. Tannen (ed) *Analyzing Discourse: Text and talk*, Georgetown University Press, Washington DC, 71-93, 1981.

11. Hart, J. 't, Collier, R., and Cohen, A., *A Perceptual Study of Intonation*, Cambridge University Press, Cambridge, 1990.