The intonation of backchannels in Italian task-oriented dialogues: cues to turn-taking dynamics, information status and speaker’s attitude

Michelina Savino
Dept. of Psychology and Educational Sciences, University of Bari
Piazza Umberto I, 1, 70121 Bari, ITALY
m.savino@psico.uniba.it

Abstract
One of the persisting problems in human-machine interaction is the development of voice-based dialogue systems which can really reproduce the naturalness of interaction between speakers. For achieving this goal, an important aspect of human communication worth modelling is represented by backchannels. They are lexical and non-lexical tokens giving a number of feedback during interaction, where intonation plays a crucial role. Intonational features of backchanneling may vary according to language, culture and communicative contexts. This paper offers a further contribution to the intonational modelling of backchannels in Italian, useful for improving naturalness in voice-based dialogue systems for this language. Results of pragmatic and intonation analysis of a five dialogues Map Task corpus show that backchannels can convey the intention of giving vs taking the floor by using a rising vs falling terminal contour. They also indicate that this general tendency might change when other pragmatic functions and/or paralinguistic meanings are added. In all these cases, a wide range of F0 patterns are available in Italian for cueing not only turn-taking dynamics, but also information status, discourse structure and speaker’s attitude.

Keywords: human-machine interaction, backchannels, intonation, turn-taking, paralinguistics, Italian language

1. Introduction

One of the still persisting problems in man-machine interaction is the possibility of developing voice-based dialogue systems which can really reproduce the naturalness of interaction between human beings. An important aspect of human speech communication whose modelling and implementation can strongly contribute to improve naturalness is represented by the use of backchannels or acknowledge tokens (Cathcart et al., 2003, Gravano and Hirschberg 2009, Misu et al. 2011). They include lexical and non-lexical tokens (like yes, yeah, mm, uh uh, and so on) which can be generally used for signalling that the listener is attending to the speaker and prompting her/him to go on, but the range of their possible paralinguistic meanings and pragmatic functions can be wider (understanding, agreement, appreciation, assessment, passive reciprocity, incipient speakership, etc. as mainly described in Conversation Analysis works, see for example Schegloff, 1982, Jefferson, 1983), also depending on communicative contexts. Since most of backchannels are non-lexical tokens, a crucial role in signalling those functions is played by intonation, where intonational features can be obviously language-dependent. In fact, recent studies have described the main prosodic characteristics of acknowledge tokens, also for modelling purposes in man-machine interaction, in a number of languages (for example Caspers 2000 for Dutch, Jurafsky et al. 1998, Benus et al. 2007 for American English). As far as Italian is concerned, one past study is devoted to a preliminary investigation on duration and tonal features of a wide range of short expressions which include backchannels (Cerrato and D’Imperio, 2003). In all these studies, analysis was carried out on task-oriented dialogues.

Aim of this paper is providing a further contribution to describe the role of intonation in Italian backchannels, whose modelling can be useful for improving naturalness in Italian voice-based dialogue systems. Analysis is based on task-oriented dialogues, making results comparable with those obtained for other languages. Also, in eliciting data some parameters for controlling and enhancing backchanneling have been used, and a more “context-based” approach in interpreting backchannel functions has been adopted.

2. Materials and methodology

2.1. Corpus

Spoken materials analysed consist of five Bari Italian dialogues elicited with a modified version of the Map Task methodology (Anderson et al., 1991), each having an average duration of 10-20 minutes, corresponding to the spoken productions of ten speakers. In a Map Task session, pairs of participants – an Instruction Giver (henceforth IG), and an Instruction Follower (henceforth IF) – is given a map. One of the two maps has a route drawn on it, and the task consists in reproducing as accurately as possible the route on the other map by exchanging information via the verbal channel. The task is complicated by the fact that the two maps are not identical in terms of presence and position of the landmarks, thus stimulating possible misunderstanding like in natural, everyday interaction.

Differently from the original Map Task method (Anderson et al, 1991), in Bari Italian sessions participants were not informed in advance that the two maps were different; neither they were told that the maps were identical, even though this is was they assumed (Grice and Savino, 2003). In other words, before starting the task participants assumed they initially shared the same background knowledge (how this aspect has an influence on backchannel intonation will be discussed in section 3.2).

In each recording session, eye contact was always inhibited, in order to maximise the use of the verbal channel for communication, including backchannels. In
fact, it has been observed that in Map Tasks lack of eye contact stimulates the production of verbal backchannels in order to provide dialogue partners with increased verbal feedback (Boyle et al, 1994). Because of the impossibility of seeing each other, participants need to use more verbal feedback also for regulating turn-taking as much efficiently as possible.

It can be noted that task-oriented dialogues like Map Tasks, including the possibility of controlling some parameters like eye contact, are particularly suitable for studying and modelling intonation of backchannels for human-machine interaction, especially for voice-based dialogue systems. In fact, in interactional contexts successfullness of information transferring is crucial for achieving the (communicative) goal, where verbal feedback and turn-taking regulation play a crucial role in making information exchange successful.

2.2. Pragmatic and intonation analysis
As a first step, pragmatic analysis of the dialogues based on both orthographic transcripts and audio files was carried out. Such analysis consists of pragmatic annotation of utterances in terms of conversational moves, according to the Map Task coding scheme (Carletta et al, 1997). This scheme includes a broad category for describing backchannel phenomena, namely the Acknowledge conversational move, defined as “[...] a verbal response that minimally shows that the speaker has heard the move to which it responds, and often demonstrates that the move was understood and accepted” (Carletta et al, 1997: 19).

Since we were also interested in looking at possibly specific intonational strategies used in backchannels for regulating turn-taking, in the pragmatic analysis we included a further distinction (introduced by Jefferson, 1983 and used later by Jurafsky et al, 1998) between:

- Acknowledge tokens reflecting Passive Recipieny (henceforth PR), also called continuers, acknowledging that the other speaker still has the turn;
- Acknowledge tokens reflecting Incipient Speakership (henceforth IS), indicating intention to take the floor, reflecting “[...] preparedness to shift from recipiency to speakership” (Jefferson, 1983: 4).

Decisions on whether an Acknowledge token could be classified as an example of PR or IS were based on whether or not a change of speaker occurred after that token. Following Cathcart et al. (2003), we identified Transition Relevance Places1 (henceforth TRPs) (Sacks et al, 1974) at move boundaries. When a change of speaker was observed after the Acknowledge token/move, that token was categories as an ACKNOWLEDGE-PR, like in the following example:

IG: now turn around the lake
    INSTRUCT
IF: mm
    ACKNOWLEDGE-PR
IG: and then go towards the animal house
    INSTRUCT

When after the Acknowledge token/move the same speaker went on speaking (i.e., no change of speaker was involved), that token was labelled as an ACKNOWLEDGE-IS, like in the following example (bars indicate move boundaries within the same turn):

IG: once reached the shop, go down towards the bar
    INSTRUCT
IF: mm || shall I go from the left or from the right side?
    ACKNOWLEDGE-IS || QUERY-W
IG: from the left
    REPLY-W

This simple criterion was adopted in order to avoid circularity in the interpretation of backchannels, i.e. to avoid using intonation for classifying acknowledge tokens as reflecting PR or IS.

Acknowledge tokens selected for intonation analysis are all lexical and non-lexical (monosyllabic and bisyllabic) tokens used in Acknowledge moves, such as si (yes), mm, okay (this token is used in English by Italian speakers), eh, aha, etc., for a total amount of 463 tokens. All tokens have been intonationally analysed by global F0 shapes (fall, rise, fall-rise, etc.), using the Praat software tool for speech analysis (Boersma & Weenink, 1999). Results shown and discussed in this paper refer to the mostly occurring acknowledge tokens in the Bari Italian dialogues, namely si, mm, okay, eh (342 tokens). Note that these tokens are also found in the dialogues as positive replies (REPLY-Y moves) to yes-no questions (QUERY-YN, CHECK and ALIGN moves). An intonational characterisation of these tokens in relation to the two different pragmatic functions (acknowledges vs positive replies) is described in Savino (2010).

3. Results and discussion
3.1. Backchannels and turn-taking
In modelling backchannels and their relation to turn-taking, it can also be useful to determine whether token type choice is a parameter involved in such dynamics, as suggested by Conversational Analysis (for example, Jefferson, 1983). Figure 1 shows the distribution (in percentage) of the four main token types (si, mm, okay, eh) with respect to the Passive Recipiency vs Incipient Speakership features conveyed in backchanneling. To the extent of the statistical significance of our data, results show that in our dialogues Italian tokens si and mm are mostly used for reflecting PR in backchanneling (i.e. when speakers do not take the turn afterwards), token eh is mostly correlated with Incipient Speakership (i.e. when speakers takes the floor after backchanneling), whereas okay seems to be almost equally used in both cases.

These results suggest for Italian a different preference in Acknowledge token choice in relationship to PR vs IS in comparison to English. According to Jefferson (1983), in fact, in English mm is mostly used with PR, whereas yes/yeah is mainly associated with IS.

More statistical data is needed to confirm this preliminary outcome. However, it indicates that token choice is an important language and cultural parameter to take into account when modelling backchannels, also for human-machine interaction.

---

1 TRPs are defined as points for potential turn switching between conversational partners.
As to the intonation analysis, Figure 2 shows the overall distribution (in percentage) of intonation contours with a rising vs falling terminal across the two main Acknowledge token types, i.e. those implying PR (ACK_ChangeofSpeaker) and those reflecting IS (ACK_NOChangeofSpeaker). It can be observed that tokens involving taking the floor after backchanneling are predominantly characterised by a falling terminal contour, whereas the F0 shape of tokens involving not taking the floor after backchanneling ends mostly (almost 60%) with a rise. Therefore, these results suggest a specific role of intonation in signalling turn-taking during backchanneling.

![Graph showing distribution of intonation contours](image1)

**Fig. 1:** Distribution of the 4 main token types with respect to whether the backchannel token used implies Passive Reciprocation (ACK_ChangeofSpeaker) or Incipient Speakership (ACK_NOChangeofSpeaker)

Results also show that an amount (around 30%) of Acknowledge tokens with PR are characterised by falling terminal contour instead. Since Acknowledge tokens involving PR seem to show some more variability with respect to Acknowledge tokens implying IS, for the former Acknowledge token type a description of each intonation contour shape, along with a possible pragmatic interpretation of different intonation choices derived by looking at their production in specific dialogue contexts are discussed in more details in the following section.

![Graph showing distribution of intonation contours](image2)

**Fig. 2:** Distribution of rising vs falling terminal contours across the two token types, i.e. Acknowledge_PR (ChangeofSpeaker) and Acknowledge_IS (NoChangeofSpeaker)

3.2. Backchannels, information status and speaker's attitude

As illustrated above, Italian acknowledge tokens implying PR are mostly characterised by a terminal rising pitch. As shown in Figure 3, for F0 rising backchannel type a wide range of contour choice is available: rise, fall-rise, stylised rise, and high rise (in the figure, rise and high rise are collapsed in one category).

In the dialogues, it is observed that rise, fall-rise and stylised-rise contours are normally found in tokens signalling understanding/agreement about information (typically, instructions) being received by the dialogue partner.

The rising contour seems to be the “default” F0 shape for this token type. This outcome seems to reflect what is generally attested in the literature as typical of continuers also in other languages. An example of token *si* with a rising contour is shown in Figure 3(a).

![Graph showing distribution of intonation contours](image3)

**Fig. 3:** Distribution of intonation contours across the 4 Acknowledge token types implying Passive Reciprocation (i.e. when speaker does not take the floor after acknowledging)

It is also observed that the fall-rise and the stylised rise are used by participants when information – typically instructions about presence and/or position of landmarks – are fully compatible with features on her/his own map. In other words, they seem to be used for confirming the current status of common background knowledge, and for conveying a consequently positive speaker’s attitude.

In particular, the stylised rise is only found in *mm* tokens, and in this case they are produced as bisyllabic, as also observed in English backchannels (Ward, 2004). The stylised rise is realised as a pitch step-up from the first to the second syllable, where the pitch excursion is not wide.

An example of a *mm* token with stylised rise is shown in Figure 4(b), whereas Figure 4(a) offers an example of *mm* with fall-rise (in this case the token is realised as bimoraic).

![Graphs showing intonation patterns](image4)

**Fig. 4:** Acknowledge_PR *mm* realised as a fall-rise (a), and as a stylised rise (b). Note that in the first case the token is realised as bimoraic, in the second as bisyllabic.
A number of backchannels with a high rising pitch contour were also encountered, typically produced by instructions. Followers at around the end of a set of instructions for completing a (sub)task. This contour type has been already described in Bari Italian for signalling pre-finality, i.e. marking the antepenultimate item in a sequence (Savino, 2001; 2004). This suggests that in acknowledge tokens with PR this contour conveys speaker’s understanding/agreement that the end of a (sub)set of instructions for completing a (sub)task is approaching. An example of such a contour is shown in Figure 5(b).

![Fig. 5: Acknowledge_PR si realised with a rise (a) and with a high rise (b), in this second case for pre-signalling that the end of a (sub)task is approaching.](image-url)

As to the Acknowledge token with PR realised with a falling terminal, our intonation analysis reports cases of fall and stylised fall melodic shapes (see distribution in Figure 3). By looking at the dialogue contexts where a falling rather than the typical rising terminal contour is encountered, it is observed that fall and stylised fall F0 patterns are usually found where participants have not discovered yet that the two maps are different. Such contours are typically produced in backchannels by IFs when receiving instructions for drawing the route which are not compatible with presence and/or position of landmarks on her/his own map. Even though the use of backchannels generally implies signalling understanding and agreement, in these cases a falling contour seems to convey disagreement or disappointment about what is assumed to be the currently shared background knowledge. Therefore, a falling contour seems to signal a negative backchanneling, as there is a mismatch between participants’ assumed common background knowledge at that moment of the interaction.

Some cases of fall and stylised fall in acknowledge tokens with RP are also found at the ending turn(s) of the dialogue. Because of their position within the dialogue, and the melodic shape typically associated the meaning of “finality”, these backchannels can be interpreted as signalling understanding/agreement that the task has been completed. Some other cases of such falling contours are found in tokens acknowledging interlocutor’s preceding acknowledgement. This outcome is particularly interesting, as the occurrence of this pragmatic type of backchannels contrasts with the formal definition of these token types proposed by Ward & Tsukahara (2000). According to these authors, in fact, backchannels do not require acknowledgement by the other speaker. On the contrary, our results show that such a pragmatic circumstance is possible, demonstrating that backchannel strategies can largely vary, depending on individual, cultural and above all communicative contexts variables.

Examples of a mm token realised with a fall (a) and okay with a styled fall (b) are shown in Figure 6. Note that the styled falls are intonationally realised as an F0 step-down from the first to the second syllable, with a relatively reduced pitch excursion.

![Fig. 6: Acknowledge_PR mm with falling contour (a), and okay with styled fall (b). The first observed when speaker discovers a mismatch in assumed shared background knowledge, the second at the end of the task.](image-url)

<table>
<thead>
<tr>
<th>Pragmatics/Paralinguistics</th>
<th>rise</th>
<th>high rise</th>
<th>fall-rise</th>
<th>styl-rise</th>
<th>fall</th>
<th>styl-fall</th>
</tr>
</thead>
<tbody>
<tr>
<td>acknowledges understanding/agreement about received instr/info (&quot;default&quot;)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>confirms current status of assumed common BK</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>signals that end of set of instructions for completing a (sub)task is approaching</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>signals disagreement or disappointment about what assumed to be the currently shared BK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>acknowledges interlocutor’s preceding acknowledgement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>signals end of (set of instructions for completing) the (sub)task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Intonation contour and their association with pragmatic functions/paralinguistic meanings in acknowledge tokens with PR as observed in Bari Italian Map Task dialogues (BK = Background Knowledge)

Finally, the distribution of intonation contours across the four main acknowledge tokens with PR (Figure 3) shows that the token eh is always realised with a falling F0 pattern. In this case, there is a specific choice not only
in terms of melodic shape, but also in terms of token type for cueing information status, speaker’s attitude and discourse structure while acknowledging understanding, as described above for falling terminal backchannels.

The intonational choices of acknowledge tokens in relation to the pragmatic functions and paralinguistic meanings discussed above are schematised in Table 1.

4. Conclusions

Results from pragmatic and intonation analysis of Bari Italian Map Task dialogues have pointed to a number of intonational features for modelling backchannels which can be useful for improving naturalness in speech dialogue systems for Italian.

Results have shown that when speakers do not take the floor after backchanneling, they produced acknowledge tokens predominantly characterised by a rising terminal intonation pattern. On the other hand, when speakers take the turn right after backchanneling, their acknowledge tokens have a falling terminal instead. Therefore, in Italian the intention of taking the floor or not while backchanneling seems to be conveyed intonationally by a falling vs rising terminal F0 contour.

Specific token type choice might also play a role.

Outcomes of our analysis also suggest that intonation in backchannels is important not only for signalling turn-taking, but also for cueing discourse structure and further paralinguistic/pragmatic meanings like the status of participants’ assumed shared background information at that time of interaction, and speakers’ attitude. In these cases, the general observed “rule” of a terminal rising for signalling the intention of yielding the floor might not be applied.

Moreover, a wide range of F0 shapes are available for cueing all these paralinguistic/pragmatic meanings.

Future work will include carrying out perceptual experiments for confirming the above mentioned paralinguistic additional meanings in backchannels, along the line, for example, of Stockmeier et al. (2007).

References


