Proceedings of DiSS-LPSS Joint Workshop 2010,
25-26 September 2010, Tokyo, Japan, pp. 103-106.

Hesitation and uncertainty as feedback

Kristiina Jokinen

Department of Speech Sciences, University of Helsinki, Finland
Kristiina.Jokinen@helsinki.fi

Abstract

This paper deals with the signals that are used to express hesitation and uncertainty in conversational interactions. It studies the relation between gesturing, body posture, facial expressions, and speech, and draws conclusions of their role and function in the interpretation and coordination of interaction with respect to the basic enablements of communication. Dialogues are assumed to be cooperative activity that is constrained by the participants' roles, social obligations, and communicative situation.

Index Terms: hesitation, uncertainty, interaction, speech

1. Introduction

In spoken dialogue management, one of the important techniques for smooth interaction concern feedback through which the speakers provide information about how their communication is proceeding. Feedback is one of the most important cohesion devices in human conversation, aiming at grounding information and signalling the failure or success of the listener's processing of a speaker's utterance [1]. In spoken dialogue systems, it is common to study conversational feedback and grounding processes to build models for effective responses. Feedback models concern positive or negative evaluation of the presented information as well as updating the existing background information for the purpose of achieving task-related goals and building shared information and mutual rapport.

Much of the research is directed towards positive feedback that allows the partners to build the shared context by adding information that is considered helpful and relevant. The starting point has usually been an explicit indication of the participants' level of understanding, i.e. explicit verbal utterances that indicate the user's intake of the information. Models of feedback and grounding such as [2] and [3] include special grounding acts that took care of the level of grounding, while [4] and [5] focus on the basic enablements and cooperative principles of interaction in order to model feedback. Characteristics of speech in grounding and backchannelling were studied e.g. in [6], [7], [8]. Feedback is also closely related to turn-taking, and in the most recent work the focus has been on flexible turn management: [9] introduces an importance-driven bidding approach to turn-taking, while [10] presents an incremental dialogue management model. The goal is to avoid the traditional fixed turn alternation between the system and the user, and to make asynchronous language interpretation and production available.

Less attention has been paid on the speaker's hesitation and uncertainty concerning a particular topic or piece of information. Hesitation provides negative feedback in that the speaker conveys his uncertainty, doubt, and lack of knowledge to the partner, and it is often regarded as a signal of conversational disfluency. From the point of view of speech technology, hesitations and uncertainty expressions pose challenges for speech recognition and spoken interaction, and emotional and expressive speech should be recognized and interpreted correctly in order to avoid clumsy, misinterpreted situations [11]. Moreover, in multicultural contexts speech-based interaction requires studies about the different strategies that the agents use to express their understanding or hesitation given their different cultural backgrounds [12, 13].

Besides expressing the factual information, the speakers also indicate their commitment and attitudes to the content of their utterances. Commitment is not a binary-valued relation between the speaker and the utterance content but a continuum from the interlocutors' strong intentions to their benevolent or even ignorant attitudes. For instance, shared understanding and mutual bonds are "tacitly" constructed through the pieces of information that display the speaker's attitude to the presented information: agreement and similar view-points tend to increase affection between the partners, while the opposite ones lead to disconcerting, even hostile situations. Negative attitudes may thus be expressed through hesitation rather than straightforwardly uttering negative expressions, and unwillingness to commit oneself to a particular issue can be signalled by overt uncertainty so as to soften the impact and to conform to politeness codes and face-saving strategies.

In this paper we discuss hesitation and hesitation related phenomena. Hesitation expressions can be divided into two types: those that express the speaker's uncertainty concerning the truth-value of the content (factual hesitation) and those that express the speaker's uncertainty concerning the relevance or importance of the content (evaluative hesitation). They have different manifestations in natural interactions: the former is usually explicitly verbalised as the speakers are aware of the limits of their knowledge and wish to avoid accusations of providing possibly false information, whereas the latter is signalled through non-verbal signs and governed by social politeness codes, which can cause misunderstandings, especially in intercultural communication, if the speakers are not familiar with the communicative functions of the signals.

We focus on how hesitation is expressed in social situations and point out that it can also function as a means to coordinate and control the dialogue. Methodologically the research relies on observation and experimentation: human-level observations and analysis of dialogue events are combined with signal-level measurements of gaze, gesture, and speech. Conclusions are drawn with respect to semantic themes of gestures and the coordinating function of hesitation expressions in communicative situations.

2. Previous research

Hesitation is characterized by lack of knowledge and its marking in speech is varied. In English, it ranges from different types of hesitation markers (*uhh, umm*) to pauses (silence) and slow speaking rate. Fundamental frequency F0 is also shown to rise before pauses that occur in major syntactic boundaries, but not if the pause occurs elsewhere. It has also been shown that perception of hesitation in speech is strongly influenced by deviations from an expected temporal pattern, and that different syntactic conditions have an effect on how much changes in prosodic features contribute to the perception of hesitation [14]. In Japanese, similar characteristics are also found. Different types of backchannels can be used, varying their commitment to presented information from non-committal "I'm listening" signals to more serious agreement sounds. For instance, [8] found that the prosodic and temporal features of a response carry information about how the speaker has grounded the information expressed in the previous utterance by the partner, especially if the speaker repeats a portion of the utterance. The features include the delay in responding, pitch and tempo, and boundary tone: longer delays, higher pitch, slower tempo, and rising boundary tone signal higher integration. A typical feature of Japanese is backchannelling which signals the status of the conversation and its emotional effect to the partner [6]. Studies concern prosodic and syntactic features of backchannels [7], while [15] predicted utterance complexity based on filled pauses.

Considering gesturing and hesitation, [16] shows that gestures and speech are closely linked, and that the gesture phrases, defined in terms of the perceptually marked preparation, stroke, and recovery phases of the action, are related to spoken phrases (intonation units). Kendon further identified different gesture families which have their own semantic theme such as stopping or halting of an action, or offering and giving of ideas and concepts. Although the themes can be considered universal interpretations of hand gestures, gesturing in general is a culture dependent means to communicate. For instance [12] noticed that a shoulder shrug can be interpreted in two opposite ways: it can be a sign of hesitation and uncertainty, or of self-confidence and certainty. Assuming that the semantic theme of a shoulder shrug is associated to non-continuation of communication, the speakers obviously focus on different reasons for the non-continuation: either on the lack of ability to continue (uncertainty) or the lack of willingness to continue (confidence). It is thus useful to compare and categorize feedback strategies in intercultural contexts, so as to fully understand the interlocutor's expectations about the appropriate behaviour in creating shared understanding.

3. Examples of hesitation

In this section examples of hesitation behaviour are given. Assuming that the basic enablements to continue the dialogue (contact, perception, understanding; see [17]) are not fulfilled, we can distinguish the following situations:

1) lack of own ability to continue (knowledge, skills)
2) lack of permission to continue (situational issues)
3) lack of willingness to continue (attitude).

The first one concerns the speaker's own inability to continue: either the speaker lacks necessary knowledge or the necessary skills to perform the actions. This seems to be the most straightforward interpretation and is often manifested by the speaker trying to find appropriate words or phrases. It can also be expressed verbally by utterances that indicate the speaker's level of commitment to the information *(I'm not sure, I assume)*. The second category concerns situations where the speaker is prevented from continuing by contextual reasons: there are social, situational, or environmental constraints that affect the speaker's behaviour. These reasons are usually culturally specified and learnt through interaction. The hesitation can be marked by longer pauses which indicate the speaker's wish to change the topic and allow the partner to take the turn. The third category relates to the speaker's own intentions and attitudes: the speaker does not want to continue. Socially this can be the most problematic class since the speaker refuses to continue on the introduced topic and can thus create a conflict which needs to be resolved.

Below examples of these cases are presented, focusing on the relation between speech and gestures in the interpretation. The examples are taken from the conversational multiparty corpus collected at ATR/NICT in Japan [11]. The data contains free-flowing conversations among four participants and is about 1,5 hours long. The dialogues are conducted in English, and topics vary from casual chatting and story telling to travel information and cultural conventions.

Figure 1 shows snapshots of the speaker who expresses his lack of knowledge verbally and by gesturing:

they would use the…yeah…to show that you're angry to make a distance… [*probably…*] *(1.4sec)* *I don't* [*know…*]
[left hand over left shoulder] [eye-brows]

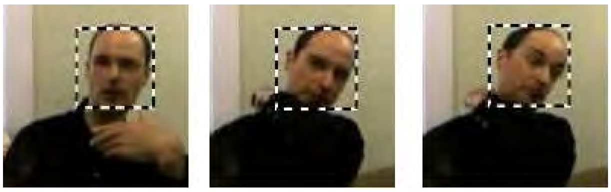


Figure 1. Snapshots of the speaker lifting his hand and raising eye-brows while uttering *to make a distance… probably…… I don't know…*

The interlocutors talk about the use of surname to address people: in some cultures, it is common to use surname also when addressing one's friends and colleagues, while in other cultures it signals social distance and the speakers can effectively use it to distance themselves from the others. When uttering the hesitation word *probably*, the current speaker simultaneously lifts his left hand up over the right shoulder as if scratching his back, and, after a pause, explicitly expresses his lack of knowledge to continue and also emphasises the word *know* by raising his eye-brows while still keeping his hand raised up and back. The hand gesture can be related to real itching of one's back, but it also seems to have a communicative meaning in that it stops the conversation and draws the partners' attention to the tentative hypothesis made earlier about the use of surnames. The gesture also withdraws the speaker's hand from the space immediately in front of him, as if iconically displaying the distancing effect of the use of surname. The semantic theme of the gesture seems to accord with that of Kendon's Open Hand Prone family, as it expresses in general halting of an action.

The whole phrase *I don't know* starts after a fairly long pause (1.4 seconds), and is speeded up so that it lasts only about 0.5 seconds. Speeding up is common if one wants to finish the turn, and the speaker indeed seems to want to terminate his turn: the utterance simply states the reason for this, lack of knowledge, and is accompanied by raising one's eye-brows. Usually the raising of one's eye-brows has interrogative interpretation, i.e. it is a sign of surprise, astonishment, or confusion, and in this case, the expression may well invite the partners to contribute to the discussion. In fact, the partners could have taken the turn already during the pause, but as they did not, the raising of one’s eye-brows can be seen as an emphasis of the communicative choice to invite the partner to speak. A partner immediately understands this and continues by saying: *I remember at school we had a teacher who called us by surname; that was really strange.*

The next example ((Figure 2)) occurs later in the same session: the participants discuss the use of Japanese endings *san, sensei, kun*, which carry meanings related to social hierarchy, in connection with the person’s first name.

but the problem is, if you…if you say… hmm…
if you [*use the first name,...*] *then you*
[vertical hand palm open down]
[*cannot add san or any respect to that* …[*I … I think*
[vertical hand sideways [curled hand to chin

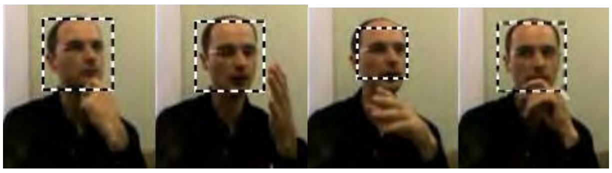


Figure 2. Snapshots of the speaker uttering *But the problem is if you -- use the first name – cannot add san or any respect -- I I think*

In this example the speaker’s verbal and non-verbal behaviour seem to contradict each other. The speaker expresses hesitation verbally, but also seems confident of what he intends to say. The speaker’s utterance is retarded, it contains long pauses and at the end also stuttering *I-I think,* but his body language is bold and confident. The speaker looks at his partners and his accurate hand beating structures the message: the first beat down fixes *use of the first name* as the topic, while the open palm sideways paints the alternatives *san or any respect*. The final gesture on the chin is the same as in the beginning of the turn where the speaker’s first expresses his ponderings upon possible problems; it iconically presents the speaker’s returning back to the start state after delivering the contribution. The verbal presentation resembles hesitation (long pauses, verbal hesitation markers, etc.) but non-verbal behaviour gives an impression that the speaker holds the floor and has the right to continue to present his thoughts. In fact, it can be hypothesized that the speaker has learnt the non-verbal behaviour in social situations so as to keep his turn while thinking and planning his expression. The speaker is rather confident in that he intends to say something but since long pausing and thinking are neither desirable nor acceptable in socially fluent conversations, we can say that there is a social constraint not to continue. The hesitative verbal behaviour, which is part of the speaker’s natural planning, is in this example accompanied by non-verbal gesturing that the speaker has learnt as a successful strategy to keep the turn and to emphasise the importance or relevance of what one is saying.

The last example in Figure 3 concerns the speaker’s own communication management and can be understood as signalling that the speaker is not willing to continue the topic. The discussion has dealt with strange things people do when fighting, and one of the partners has asked what the speaker meant by saying “it could be worse”. The answer is accompanied by a complicated set of hand gesturing that first emphasises the speaker’s view of the stupidity of the fighting, and later his lack of knowledge and willingness to continue with the topic as there is nothing more to say about. The discussion draws to a halt with a silence about four seconds.

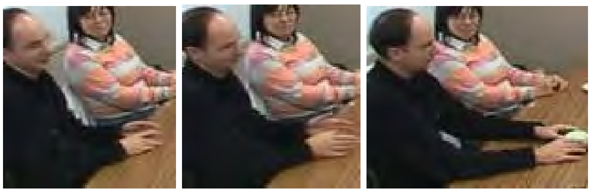


Figure 3. Snapshots of the speaker and his gesturing: *each other but…[but…no... I don’t know] …very stupid*

Well I mean worse than…what you could possibly imagine
… when people fight they manage to get the maximum for
[each other [but [no…
[hands beat [open palms sideways [flick back palm down
[I don’t know
[palm open sideways then back palm down
[it’s just very stupid
[hands reach coffee cup in front, gaze down

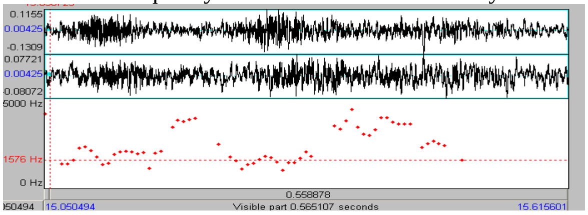
4. Data analysis

The examples show that, while the speaker is uncertain or unable to continue the conversation, his interaction management can be quite different and be reflected in the simultaneous non-verbal communication. The speaker may lack ability, social permission, or own willingness to continue, and signal these aspects by non-verbal communication. The speaker expresses hesitation and uncertainty about how to continue depending on whether the basic enablements of communication are fulfilled or not.

In the first example, the gesture iconically presents the content of the utterance (distancing from the interaction centre), and the facial expression with the raised eye-brows invites the partners to contribute to the shared knowledge by providing their understanding of the topic. In this case, hesitation is related to the speaker’s lack of knowledge or self-confidence about the particular claim. In the second example, the speaker reckons he has the “knowledge” but may be socially constrained to give the turn to the partners while thinking of how to express his message. He signals his willingness to continue by indexical hand gestures which accurately structure the topic and “allow” the partners to take the turn at the end of the utterance when the hand returns back to the original thinking position. The final example is related to the speaker’s attitudes and lack of willingness to continue: the topic has been long and exhaustively discussed, and there is nothing more to be said. The speaker’s attitude is conveyed by his gesturing. The iconic fist for fighting is widened to an open palm which moves sideways and then returns back to the table palm down, and finally, with a strong expression of disgust, the arms are iconically extended forward as if pushing an annoying event away. Following [16], the flicking of hands palm open and sideways can be linked to offering ideas and concepts, while returning them back stops the action and metaphorically also closes the topic. The final extension of hands is a fairly clear non-verbal signal of one’s views, and conveys the speaker’s intention to finish the topic.

The types of gesturing seem to support interaction management: verbal expressions of hesitation and uncertainty are accompanied by particular gesturing that guides the partners to understand the underlying reason for the hesitation and uncertainty. The precise interpretation of the non-verbal signals is of course not possible, but it is possible to talk about semantic themes of gestures [16]. It is interesting that the partners tend to observe and interpret the signals in the intended way: the situations continue so that the partners produce their own contribution to the topic, thus cooperating with the speaker to build a shared understanding, or listen to the speaker without interrupting his turn, or provide accommodating feedback and start a completely new topic.

The speech analysis indicates that the sample utterances corroborate the earlier analyses of hesitation expressions. As an example, Praat analyses of the utterance: *problem is if you - use the first name – cannot add san or any respect – I I think* are shown in Figure 4 (variation of F0) and Figure 5 (pitch contour and intensity). For instance, the speaker does not raise his voice, but does lengthen the function word so that the speaking rate seems to slow down. The pauses occur in the middle of grammatical phrases (*if you - use…*) signalling unfinished thoughts, while repetition of pronouns (*I I think)* explicitly expresses that the speaker is uncertain about the content. Analysis also shows that the pitch contour of the last part of the utterance stays rather flat as is expected, while fundamental frequency shows a downward tendency.



4. F0 analysis of the utterance: *I I think*

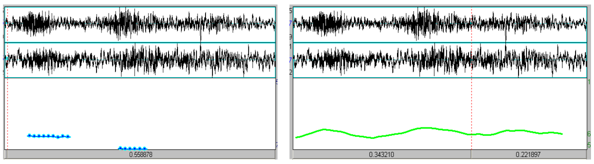


Figure 5. Pitch contour and intensity analysis of the utterance: *I I think*

5. Conclusions

Success of interaction depends on the cognitive and emotional impact of the response on the hearer, and research issues in this respect concern the question how to construct and update knowledge on the particular new information. Hesitation is an important means to give feedback about the speaker's ability and willingness to continue conversation. However, it is related to potentially dangerous situations where the speakers may lose face, and it is thus important to understand how this kind of negative feedback operates in communication and can support and also break smooth flow.

In this paper different hesitation phenomena were discussed from the view-point of the basic enablements of communication, and the main distinction was drawn between the speaker's ability, permission, and willingness to continue. Moreover, verbal hesitation and uncertainty expressions were associated with hand gestures and body posture, and tentative hypotheses of the role and correlation of non-verbal communication in their interpretation were made. The conclusion is that hesitation is not only a means to express uncertainty in communication (Fig. 1), but also functions in the coordination of interaction and the activity that the users are involved in (Figs.2 and 3). In particular, situational and attitudinal hesitation seems to be accompanied by larger, more accurate and specific gesturing which carries social conventions of symbolic gestures. Future challenges concern modelling the correlations between different speech parameters, gesturing, and culturally acceptable hesitation marking.

6. References

[1] Clark, H., Schaefer, E. 1989. Contributing to discourse. *Cognitive Science* 13:259-294

[2] Allen, J., Schubert, L., Ferguson, G., Heeman, P., Hwang, C., Kato, T., Light, M., Martin, N., Miller, B., Poesio, M., Traum, D. 1994. *The TRAINS Project: A Case Study in Defining a Conversational Planning Agent*. Technical Report 532, Computer Science Dept., U. Rochester.

[3] Traum. D. 1994. *A Computational Theory of Grounding in Natural Language Conversation*, Technical Report 545, Computer Science Dept., U. Rochester.

[4] Allwood, J., Nivre, J., Ahlsen, E. 1992. On the Semantics and Pragmatics of Linguistic Feedback. *Journal of Semantics*, 9, 1–26.

[5] Jokinen, K. 1996. Cooperative Response Planning in CDM: Reasoning about Communicative Strategies. In: S. LuperFoy, A. Nijholt, G. Veldhuijzen van Zanten (Eds.) *TWLT 11. Dialogue Management in Natural Language Systems*. Enschede: Universiteit Twente. pp. 159-168.

[6] Katagiri, Y., Sugito, M., Nagano-Madsen, Y. 1999. Forms and Prosodic Characteristics of Backchannels in Tokyo and Osaka Japanese. *International Congress of the Phonetic Sciences*, pp. 2411–2414.

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

[8] Shimojima, A., Katagiri, Y., Koiso, H., Swerts, M. 1999. An Experimental Study on the Informational and Grounding Functions of Prosodic Features of Japanese Echoic Responses. *Procs of ESCA Workshop on Dialogue and Prosody*, pp. 187–192.

[9] Selfridge, E., Heeman, P. 2010. Importance-Driven Turn-Bidding for Spoken Dialogue Systems. *Procs of the 48th Conference of ACL*, Uppsala Sweden,

[10] Skantze, G., Schlangen, D. 2009. Incremental dialogue processing in a micro-domain. *Procs of the 12th Conference of the EACL*, Athens, Greece. pp. 745-753.

[11] Jokinen, K., Campbell, N. 2008. Non-verbal Information Sources for Constructive Dialogue Management. *Tutorial at LREC-2008*. Marrakech, Morocco.

[12] Jokinen, K., Allwood, J. 2010. Hesitation in Intercultural Communication: Some observations on Interpreting Shoulder Shrugging. *Procs of the Workshop on Culture and Computing*, Kyoto, 2010.

[13] Endrass, B., Rehm, M., André, E. 2009. Culture-specific Communication Management for Virtual Agents. Procs of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS). Budapest, Hungary, pp. 281–288

[14] Carlson, R., K. Gustafson, E. Strangert, 2006. Modelling Hesitation for Synthesis of Spontaneous Speech. *Procs of Speech Prosody* 2006, Dresden.

[15] Watanabe, M., Hirose, K., Den, Y., Minematsu, N. 2008. Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication*, 50, 81-94.

[16] Kendon, A. 2004. *Gesture: Visible action as utterance*. Cambridge

[17] Allwood, J. 2002. Bodily Communication - Dimensions of Expression and Content. In B. Granström, D. House I. Karlsson (Eds.): *Multimodality in Language and Speech Systems*. Dordrecht: Kluwer, pp. 7-26.