

Annotating Interruption in Dyadic Human Interaction

Liu YANG, Catherine ACHARD, Catherine PELACHAUD

Institut des Systèmes Intelligents et de Robotique

Sorbonne University, 75005, Paris, France

yangl@isir.upmc.fr, catherine.pelachaud@isir.upmc.fr, catherine.achard@isir.upmc.fr

Abstract

Integrating the existing interruption and turn switch classification methods, we propose a new annotation schema to annotate different types of interruptions through timeliness, accomplishment and speech content level. We annotated the French part of NoXi corpus with the proposed structure and use these annotations to study the probability distribution and duration of each turn switch type.

Keywords: Interruption Annotation, Turn-taking, Dyadic Conversation

1. Introduction

Interruption is a phenomenon that often occurs in daily conversations and cannot be ignored when studying turn-taking during human-human interaction. Interruption is viewed as a special phenomenon in turn-taking, which plays a role in regulating the rhythm of dialogue. Many researchers interpret interruptions as evidence of power and dominance because they violate the principle of one person speaking at a time (Ferguson, 1977), (Tannen and others, 1991). However, depending on the context, some interruptions are not regarded as signs of power and discomfort, but as cooperative behaviors, which provide help to the speaker (Yang, 2001). Interruptions are therefore important during a dialogue and can have very diverse functions.

Our aim is to study interruptions in human-human interaction to understand the impact of interruption types on the on-going interaction and how the interrupter is perceived by the interruptee. To this aim, we propose a schema to annotate interruption based on theoretical models and use it to analyze a human-human dialogue corpus.

2. Background

Conversation analysis (CA) has been carried out since the 1960s with the work of Harvey Sacks. He proposed a study on the orderliness of conversation by describing its most basic structure: turn-taking (Sacks et al., 1978). Turn-taking is a fundamental aspect of the conversation since it is not possible to speak and listen at the same time. Interlocutors need to coordinate to exchange speaking turns to maintain the conversation.

Several conversation coding structures have been used for previous studies (Nakazato, 2000; Truan and Romary, 2021; Ten Bosch et al., 2004; Christodoulides and Avanzi, 2015; Jokinen et al., 2013; Enomoto et al., 2020; Heeman et al., 2006), however they didn't give details about different interruption types.

As a special embodiment of turn-taking principles, interruption is a natural phenomenon in conversations

and also an important topic in the conversational structure analysis. It reflects the communicator's attempt to take the floor without letting the speaker finish his/her utterance.

Allwood and colleagues have integrated interruption in their coding schema according its function (Allwood, 2001). However they considered interruption as only part of speech overlap and excluded interruption initiated within a pause.

Schegloff and Sacks (Schegloff and Sacks, 1973) defines three types of simultaneous speech: interruption, overlap and parenthetical comments such as backchannels. The backchannels are feedback messages that do not aim to take the speaking turn. Overlaps occur when the listener anticipates the end of the speaker's speech. There is a willingness on the part of the speaker to give up the turn. On the contrary, during interruptions, the listener takes the floor of the speaker, against his/her will (Schegloff and Sacks, 1973). To prevent confusion between overlap and interruption, Sacks and colleagues (Sacks et al., 1978) define thus 'overlap' can be defined as the listener over-anticipating the end of the current speaker (Sacks et al., 1978), resulting in an overlay between the last word or syllable of the current speaker and the first word of the listener. On the other hand, interruption is described as a violation of the current speaker's turn which overlap is not (Moerman and Sacks, 2010).

2.1. Taxonomies of interruption

In this section, we introduce two interruption classification methods. The first one is proposed by Beattie (Beattie, 1981), based on simultaneous speech and willingness to yield the floor, as shown in Figure 1.

In this taxonomy, there are three types of interruptions.

- Butting-in interruption, in which there is overlap but the listener fails to grab the turn and the speaker continues to speak.
- Simple interruption, in which there is also overlap, but the listener succeeds to get turns against the

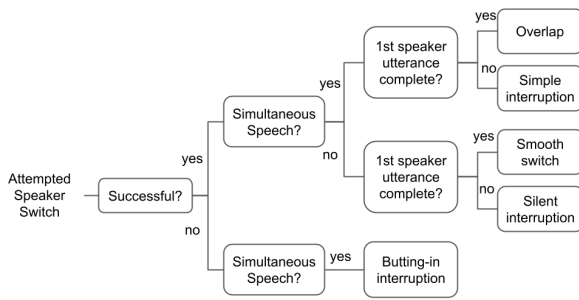


Figure 1: Classification of interruption and smooth speaker exchange (Beattie, 1981)

speaker’s wishes.

- Silent interruption, without overlap, the listener takes turns, opposing the speaker’s wishes, during a short pause.

Interruptions can also be broadly divided into two strategies along speech content: competitive and cooperative interruptions (Li, 2001).

Competitive interruption occurs when the listener interrupts to control the interaction, usually disrupting the flow of dialogue between the partners and can be seen as a conflict. A competitive interruption can be:

- Disagreement: The listener disagrees with what the current speaker is saying and expresses own opinion.
- Floor taking: The switch does not change the current topic and usually expands on the current speaker’s topic.
- Topic change: The listener changes the current topic of conversation.
- Tangentialization: The listener sums up the message from the current speaker to prevent listening to more unwanted information.

On the opposite, a cooperative interruption usually helps to maintain the conversation and can be:

- Agreement: The listener shows agreement, compliance, understanding or support to the speaker.
- Assistance: The listener interrupts to provide the current speaker with a word, a phrase or an idea to complete the utterance.
- Clarification: The listener asks the current speaker to clarify or explain the information about which the listener is not clear.

These two classification methods cover most of the situations we may encounter in conversations, but there are still exceptions, for example backchannel is not included in the first taxonomy, and some interruptions

abandoned too quickly to define their type cannot be classified using the second taxonomy. We propose a new structure to merge and complete these two methods.

3. NoXi Corpus

We choose to use NoXi corpus (Cafaro et al., 2017) for our study. NoXi is a multimodal database that contains free dyadic conversation through 45 given topics in seven languages.

In the NoXi database, participants take either the role of an expert that has knowledge on a given topic or the role of a novice that wants to learn about this topic. Both participants have been recorded during a screen-mediated interaction on a chosen topic of expertise, for the expert, and interest for the novice. Each interaction lasts about 20mn long. The video of the participants shows almost their full body except the feet. Videos of both interactants are synchronized and transcribed.

For our work, we focus on the French part of NoXi corpus, which contains 21 dyadic conversations, for about 7h in total.

4. Annotation Schema

In this section, we explain the new annotation schema, annotation method, annotation accuracy and final annotation data.

4.1. Definition

We first provide definitions of terms used in our annotation schema (Section 4.2), the '[' and ']' in the examples represent the start and the end of a simultaneous speech.

First, Smooth turn-exchanges and Interruptions are not distinguished by gaps and overlaps. A Smooth turn-exchange may occur with an overlap if the listener starts to speak overlapping several words or syllables when the speaker starts yielding the floor but has still to completed his/her utterance. In opposite, an interruption may occur with a gap when, for example, the speaker is blocked and searches for a word and the listener provides a suggestion. Thus, to find out if an interruption occurs, it is important to consider the interactants’ speech.

A successful interruption corresponds to the situation when:

- The interrupter succeeds to grab the turn and the current speaker stops talking even though s/he has not finished his/her utterance.

Example:

Novice:...basically not phy- [sical but I ...

Expert: [I agree with you for...

- The interrupter talks over the speaker (e.g. to ask a clarification question). The speaker keeps the speaking turn, but takes into account what the interrupter has said (e.g. by answering the interrupter’s question).

Table 1: Probability distribution according to the 8 types of interruption.

| Type | Agreement | Assistance | Clarification | Disagreement | Floor taking | Topic change | Tangentialization | Not identified |
|------------|-----------|------------|---------------|--------------|--------------|--------------|-------------------|----------------|
| Count | 348 | 68 | 89 | 44 | 230 | 68 | 6 | 76 |
| Percentage | 37.46% | 7.32% | 9.58% | 4.74% | 24.75% | 7.32% | 0.65% | 8.18% |

Table 2: Probability distribution of different interruption types for successful interruption.

| Type | Agreement | Assistance | Clarification | Disagreement | Floor taking | Topic change | Tangentialization |
|------------|-----------|------------|---------------|--------------|--------------|--------------|-------------------|
| Count | 310 | 64 | 84 | 42 | 188 | 66 | 5 |
| Percentage | 40.84% | 8.43% | 11.07% | 5.53% | 24.77% | 8.7% | 0.66% |

Example:

Expert: ...mushrooms and you [have to be careful. yeah, especially the optics...

Novice: [Mushrooms?]

A failed interruption corresponds to the situation when:

- The interrupter terminates the interruption before completing the utterance and let the current speaker continue his/her turn.

Example:

Expert: ...your point of view, I unders- [tand but finally] maybe it's easy ...

Novice: [Ah no no you...]

- The interrupter begins to speak and tries to get the current speaker's attention, but the current speaker does not respond to the interruption after the interrupter has completed his/her utterance and continues speaking as planned.

Example:

Novice: ...I didn't pay even one euro for Hearthtone, and I uh I still [have my meta decks up to] now, I can...

Expert: [Ah me neither I didn't pay for it.]

This last situation might be confused with a backchannel. Backchannels are short messages to show the listener's attention, but also how the listener agrees or not with, his/her attitude toward what is being said (Allwood et al., 1992). Backchannels can be longer than a single word, or even a full sentence. They differ from interruption as they are not aimed to grab the turn or expected to receive any response from the speaker. Moreover, they are not disturbing the speech flow.

4.2. Schema

In addition to the analysis of the interruption itself, we introduce other turn switches in our annotation schema. As shown in Figure 2, this schema includes three levels.

- First, we classify each voice activity change into *interruption*, *backchannel* or *smooth turn-exchange*. Interruptions correspond to the turn initiated by the listener (whether during a silence or

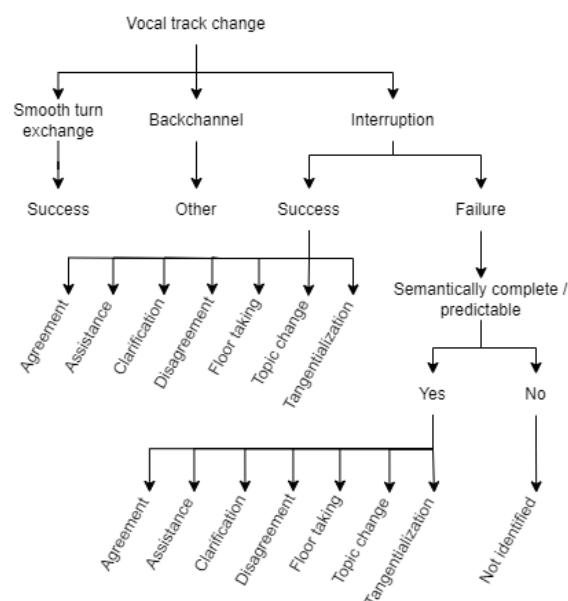


Figure 2: Interruption annotation schema

with an overlap) when the current speaker aims to continue the turn. Smooth turn exchange is the turn taken by the listener when the current speaker has finished or is about to finish and yields the turn. Backchannels are short messages indicating listener's attention, that occur during a turn.

- Then we annotate the accomplishment of turn exchange. Smooth turn exchange corresponds to successful speaking turn exchange (*success*), while interruptions can be successful (*success*) and failed (*failure*) cases. Backchannels are not grabbing the turn, the accomplishment is annotated as *Other*.
- Finally, for all the interruptions, we determine its type based on the speech content. Its type can be identified for successful interruptions. The different types we considered are listed in Figure 2. For failed interruptions, the type is determined when the speech is long enough to understand the content, otherwise it is annotated as *Not identified*.

Table 3: Probability distribution of different interruption types for Failed interruption.

| Type | Agreement | Assistance | Clarification | Disagreement | Floor taking | Topic change | Tangentialization | Not identified |
|------------|-----------|------------|---------------|--------------|--------------|--------------|-------------------|----------------|
| Count | 38 | 4 | 5 | 2 | 42 | 2 | 1 | 76 |
| Percentage | 22.35% | 2.35% | 2.94% | 1.18% | 24.7% | 1.18% | 0.59% | 44.71% |

4.3. Process

The annotation is based on the auto-detected voice activity. As mentioned earlier, NoXi contains 21 dyadic conversations, and each participant is recorded separately. To display synchronously the videos of the two participants in each conversation, and to make the two voice activities follow the video display, we use the Nova tool (Baur et al., 2020), as shown in Figure 3.

We apply voice activity detection over the whole videos. Each time, a voice activity from the current listener is detected we record its start time and annotate it according to the schema introduced in Figure 2. Then we apply a script to automatically find the end of the first Inter-Pausal Unit (IPU) that occurs after this detected voice activity. Inter-Pausal Units are widely used for conversation analysis, defined as the speech units from a single speaker without pauses (Levitan and Hirschberg, 2011). We define a pause as a silence period longer than 200ms (Demol et al., 2007) and use them to split voice activity into IPUs.

4.4. Annotation accuracy

In order to ensure the accuracy of the annotation, all videos have been annotated twice by the same annotator. Following the method proposed by (Chollet et al., 2019), there is a one-month interval between the two rounds of annotation to ensure that the annotator has forgotten the video content during the second annotation. We then extract all the annotations that are different in start point or annotation labels between both annotations. We annotate these segments a third time. After the first two rounds of annotations, we calculate the acceptance rate between them. The global annotator self-acceptance is as high as 89.5% over all the voice activity changes with same start point in the first two rounds (switch type, accomplishment and interruption type are the same in the two rounds). The self-acceptance for backchannel annotations is 93.5% over the voice activity changes marked at least once as backchannel in the first two rounds. A self-acceptance of 72.4% is obtained for interruption and 95.3% for smooth turn exchange. After the third round of annotation, we obtain a global annotator self-acceptance of 92.6% , 84.07% for interruption, 92% for smooth turn exchange and 98.8% for backchannel.

5. Statistical results

Finally we got 3983 annotation records of voice activity changes for the French part of NoXi, including 1403 smooth turn exchanges, 1651 backchannels and 929 interruptions. Thus, a change in the voice activity corresponds to a smooth turn exchange for 35% of the time,

a backchannel for 42% and an interruption for 23%.

Among all the interruptions, there were 759 (81.7%) successful interruptions and 170 (18.3%) failed interruptions.

The probability distribution according to the 8 types of interruption is given in Table 1.

Among all interruptions, cooperative interruptions account for the majority, there are 505 cooperative interruptions, being 54.36% of the total interruptions and 348 competitive interruptions, accounting for 37.46% of the total interruptions.

Subdivided into eight sub-categories, interruptions of type *agreement* are more frequent than the other types; they are the majority of cooperative interruptions. Interruptions of type *floor taking* interruptions are the most frequent competitive interruptions as can be seen in Table 1.

Combined with *accomplishment* (Tables 2 & 3), *agreement* and *floor taking* are still the two largest types in successful interruptions, but for failed ones, most interruptions could not be classified and therefore *Not identified* becomes the majority type, being 44.71% of failed interruptions.

We also analyze the length distribution of the first IPU after a voice activity change. The average duration of the first IPU after Smooth turn exchange is 4.31 seconds. Interruption takes shorter IPU with an average duration of 2.91 seconds.

When we consider the accomplishment (success/failure) of an interruption, successful interruptions have longer IPU with an average of 3.97 seconds, while failed interruptions lead to shorter IPU with an average of 2.71 seconds.

When we consider two main classes of an interruption, the IPUs following competitive interruptions (4.01 seconds on average) are 1.5 seconds longer than for cooperative ones (2.46 seconds on average).

When considering both factors, accomplishment and interruption type, the IPUs following successful competitive interruptions (4.47 seconds on average) are 1.89 seconds longer than successful cooperative ones (2.58 seconds on average).

For failed interruptions, we do not find significant difference between cooperative, competitive and not identified interruptions (Failed cooperative interruption: 1.31 seconds on average, competitive interruption: 1.08 seconds on average, and not identified interruption: 0.86 seconds on average).

6. Conclusion

In this paper we introduce a new interruption annotation schema we used to annotate the NoXi corpus. We

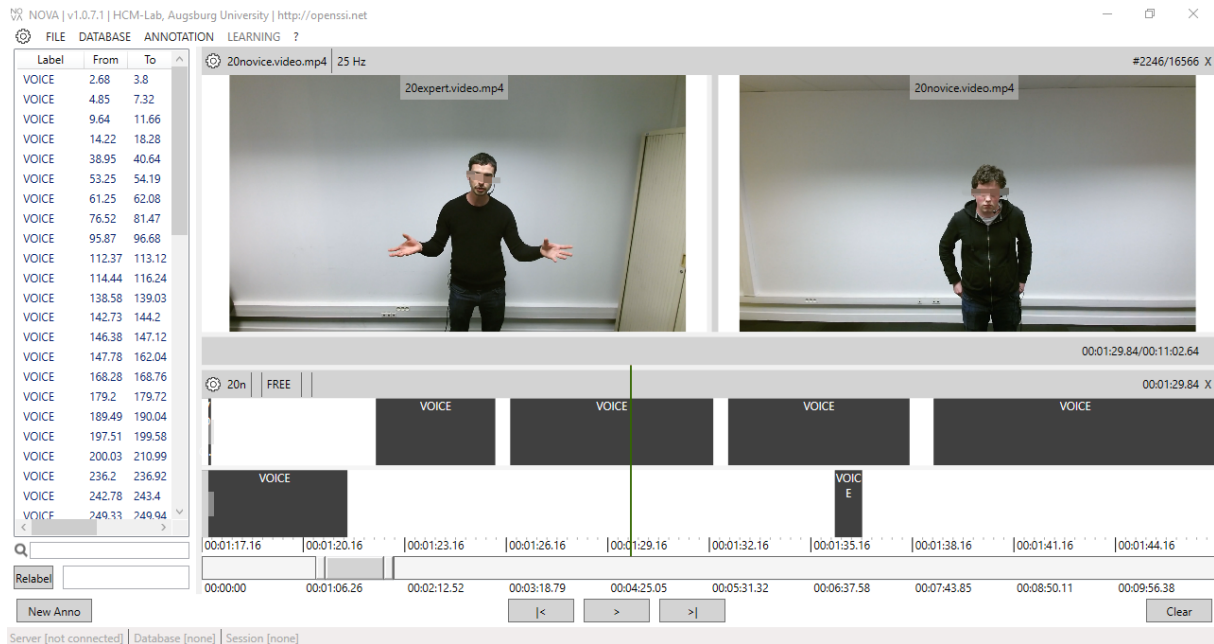


Figure 3: Nova annotation interface

conducted statistical analysis on the occurrence of the different interruption types. We also studied the length of the first IPU after the interruption and found some significant differences.

From our analysis, in the French part of the NoXi database, interruptions occur frequently in the conversations. They play an important role during conversations. Most interruptions are successfully completed and are *cooperative* interruptions. Failed interruptions are often very short, which does not allow to determine their type and justifies the introduction of the type *Not identified*. Agreement interruptions take the most part of cooperative interruptions while floor taking are predominant for competitive ones.

NoXi gathers interaction of an expert giving information to a novice on a topic that interest both of them. This particular context of interaction explains why there are more cooperative interruptions than competitive ones.

7. Acknowledgements

This work was performed as a part of ANR-JST-CREST TAPAS and ANR-JST-DFG PANORAMA project.

8. Bibliographical References

- Allwood, J., Nivre, J., and Ahlsén, E. (1992). On the semantics and pragmatics of linguistic feedback. *Journal of semantics*, 9(1):1–26.
- Allwood, J. (2001). Dialog coding-function and grammar: Göteborg coding schemas. *rapport nr.: Gothenburg Papers in Theoretical Linguistics 85*.
- Baur, T., Heimerl, A., Lingenfeller, F., Wagner, J., Valstar, M. F., Schuller, B., and André, E. (2020). ex-

plainable cooperative machine learning with NOVA. *KI - Künstliche Intelligenz*, Jan.

- Beattie, G. W. (1981). *Interruption in conversational interaction, and its relation to the sex and status of the interactants*. Walter de Gruyter, Berlin/New York Berlin, New York.
- Cafaro, A., Wagner, J., Baur, T., Dermouche, S., Torres Torres, M., Pelachaud, C., André, E., and Valstar, M. (2017). The noxi database: multimodal recordings of mediated novice-expert interactions. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 350–359.
- Chollet, M., Ochs, M., and Pelachaud, C. (2019). A methodology for the automatic extraction and generation of non-verbal signals sequences conveying interpersonal attitudes. *IEEE Trans. Affect. Comput.*, 10(4):585–598.
- Christodoulides, G. and Avanzi, M. (2015). Automatic detection and annotation of disfluencies in spoken french corpora. In *Sixteenth Annual Conference of the International Speech Communication Association*.
- Demol, M., Verhelst, W., and Verhoeve, P. (2007). The duration of speech pauses in a multilingual environment. In *Eighth Annual Conference of the International Speech Communication Association*.
- Enomoto, M., Den, Y., and Ishimoto, Y. (2020). A conversation-analytic annotation of turn-taking behavior in japanese multi-party conversation and its preliminary analysis. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 644–652.
- Ferguson, N. (1977). Simultaneous speech, interruptions and dominance. *British Journal of social and*

- clinical Psychology*, 16(4):295–302.
- Heeman, P. A., McMillin, A., and Yaruss, J. S. (2006). An annotation scheme for complex disfluencies. In *INTERSPEECH*. Citeseer.
- Jokinen, K., Furukawa, H., Nishida, M., and Yamamoto, S. (2013). Gaze and turn-taking behavior in casual conversational interactions. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 3(2):1–30.
- Levitan, R. and Hirschberg, J. B. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Interspeech*.
- Li, H. Z. (2001). Cooperative and intrusive interruptions in inter-and intracultural dyadic discourse. *Journal of language and social psychology*, 20(3):259–284.
- Moerman, M. and Sacks, H. (2010). Appendix b. on “understanding” in the analysis of natural conversation. In *Talking Culture*, pages 180–186. University of Pennsylvania Press.
- Nakazato, S. (2000). Japanese dialogue corpus of multi-level annotation. In *1st SIGdial Workshop on Discourse and Dialogue*, pages 1–8.
- Sacks, H., Schegloff, E. A., and Jefferson, G. (1978). A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction*, pages 7–55. Elsevier.
- Schegloff, E. A. and Sacks, H. (1973). *Opening up closings*. Walter de Gruyter, Berlin/New York Berlin, New York.
- Tannen, D. et al. (1991). *You just don’t understand: Women and men in conversation*. Virago London.
- Ten Bosch, L., Oostdijk, N., and De Ruiter, J. P. (2004). Turn-taking in social talk dialogues: temporal, formal and functional aspects. In *9th International Conference Speech and Computer (SPECOM’2004)*, pages 454–461.
- Truan, N. and Romary, L. (2021). Building, encoding, and annotating a corpus of parliamentary debates in xml-tei: A cross-linguistic account. *Journal of the Text Encoding Initiative*.
- Yang, L.-c. (2001). Visualizing spoken discourse: Prosodic form and discourse functions of interruptions. In *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue*.