The Distracted Ear: How Listeners Shape Conversational Dynamics

Auriane Boudin, Stéphane Rauzy, Roxane Bertrand, Magalie Ochs, Philippe Blache

The Distracted Ear: How Listeners Shape Conversational Dynamics

Auriane Boudin[1,2,3], **Stéphane Rauzy**[1,3], **Roxane Bertrand**[1,3]
Magalie Ochs[2,3], **Philippe Blache**[1,3]
[1]Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France
[2]Aix Marseille Univ, CNRS, LIS, Marseille, France
[2]Institute of Language, Communication and the Brain, Marseille, France
{firstname.lastname}@univ-amu.fr

Abstract

In the realm of human communication, feedback plays a pivotal role in shaping the dynamics of conversations. This study delves into the multifaceted relationship between listener feedback, narration quality and distraction effects. We present an analysis conducted on the SMYLE corpus, specifically enriched for this study, where 30 dyads of participants engaged in 1) face-to-face storytelling (8.2 hours) followed by 2) a free conversation (7.8 hours). The storytelling task unfolds in two conditions, where a storyteller engages with either a ”normal” or a ”distracted” listener. Examining the feedback impact on storytellers, we discover a positive correlation between the frequency of specific feedback and the narration quality in normal conditions, providing an encouraging conclusion regarding the enhancement of interaction through specific feedback in distraction-free settings. In contrast, in distracted settings, a negative correlation emerges, suggesting that increased specific feedback may disrupt narration quality, underscoring the complexity of feedback dynamics in human communication. The contribution of this paper is twofold: first presenting a new and highly enriched resource for the analysis of discourse phenomena in controlled and normal conditions; second providing new results on feedback production, its form and its consequence on the discourse quality (with direct applications in human-machine interaction).

Keywords: Multimodal Corpus, Feedback, Spontaneous Conversation

1. Introduction

Conversation is a common everyday activity, but its apparent simplicity hides an underlying complexity. Although it may appear to be an effortless exchange of words and ideas, numerous intricate linguistic and cognitive processes underpin the success of a conversation. A constant stream of multimodal information is processed at the same time as the multimodal signal is generated. Sacks et al. (1974) has shown that conversations has a structured and organized nature. The discursive role between participants frequently and quickly changes. Even in the apparently less *active* role of the listener, the act of providing feedback assumes a pivotal role in discourse construction, as highlighted by various studies (Bavelas et al., 2000; Stivers, 2008; Bertrand and Espesser, 2017). Understanding this phenomenon and more generally the structure and the dynamics of conversations is then of deep importance in many respects, in particular in the perspective of developing natural conversational agents (Poppe et al., 2011; Truong et al., 2011; Glas and Pelachaud, 2015). Such a goal requires specific datasets making it possible to conduct works for understanding the role of feedback on discourse elaboration and quality and develop multimodal conversational models.

The goal of this paper is to investigate the impact of listener disruption during conversations and its consequences on feedback production and the quality of interaction. Our work is based on (Bavelas et al., 2000) experiment, consisting in face-to-face storytelling with either an attentive or a distracted listener. Our study incorporate a free conversation part, records neuro-physiological signals, and provides comprehensive multimodal annotations for both the main speaker and the listener.

The data acquired during this experiment constitutes the SMYLE corpus (Boudin et al., 2023b), a new resource for studying the production/perception system of conversation across various dimensions: discourse perception, feedback production, common ground elaboration, information transfer, etc. We present in a first part the details of the SMYLE corpus and its large set of annotations. In a second part, we provide new results on feedback production, its form and its consequence on the discourse quality.

2. The Bavelas' Original Study and our Goals

In their seminal work, (Bavelas et al., 2000) explore the nature of social communication during face-to-face interactions by studying conversational feedback. We present in this section the

main findings of this study and new hypothesis motivating our work.

2.1. Bavelas' Experiment

Bavelas et al. (2000) argue that a listener helps the main speaker by producing different types of reactions, namely *generic feedback* and *specific feedback*. *Generic* feedback refers to a response that show understanding and invite the main speaker to continue speaking. This response is mostly conveyed through short vocalizations (*"mh mh", "ok", etc.*) and/or by nodding. The main function of *generic* feedback is to help the speaker in monitoring the interlocutor's comprehension. In contrast, *specific* feedback helps the speaker to tell a story by displaying a range of behaviors (happiness, sadness, horror, surprise, fear, etc.). Specific feedback is directly related to the content of the narrator's speech. These responses can include verbal and gestural content (wince, smiling, laughing, hand movements, head movement, etc.) or mimicry. When the listener provides specific feedback, he/she also becomes to a certain extent a co-narrator by participating actively to the discourse (Bavelas et al., 2000).

Bavelas et al. (2000), propose an original experiment to explore the role of the listener during storytelling. Thirty-four dyads of participants engaged in a storytelling task of a close-call or near-miss incident. The experiment comprises two conditions: a **control condition** asking the listener to summarize the story, and a **distracted condition**, where the listener has to count and press a button each time the main speaker uttered a word beginning with the letter *t*.

Conversations were video-recorded. A post-experimental analysis was conducted, in which third-party analysts watch the videos of the storyteller and assessed the quality of story endings according to their pace, denouement, choppiness and the narrator's attempt to justify the story's closure.

Four hypotheses were tested in their work. Hyp. #1: generic and specific feedback serve different functions. Hyp. #2: the two types of feedback occur at different stages during storytelling, with generic feedback potentially being produced early on, while specific feedback, requiring more information, appears later in the narrative. Hyp. #3: distraction has a greater impact on the production of specific feedback, as it demands a higher level of comprehension. Hyp. #4: storytelling relies on a collaborative process, with the listener playing an active role in constructing the narrative. When the listener is distracted, the quality of the story is likely to be negatively affected.

The study's findings provide support for all four hypothesis. The results indicate that specific feedback tends to occur later in the storytelling than generic feedback. Additionally, the distracted listener produced significantly fewer specific feedback than the attentive listener (0.08 specific feedback/minute in distracted condition, 2.21 specific feedback/minute in attentive condition). There was also a significant effect for generic feedback (8.91 generic feedback/minute in the normal condition, 7.44 feedback/minute in the distracted condition). Moreover, the scores from the assessment of story endings demonstrate that the word-count condition has a negative impact on the quality of the narration. Finally, no gender-based effects were observed on the feedback rate.

The distraction task was reused by (Kuhlen and Brennan, 2010), wherein participants were instructed to tell two jokes (provided by the experimenters) to either an attentive or a distracted listener. To investigate the influence of speakers' expectations on listeners' behavior, the design ensured that, in the distracted condition, half of the speakers expected interaction with a distracted listener and half did not. Similarly, in the normal condition, half of the speakers anticipated interaction with a normal listener, and the other half with a distracted listener. The distraction task employed in this study consisted of counting the occurrences of the word "*and*", which was considered less disruptive than the original task. Notably, this study revealed a decrease in feedback production within the distracted condition. Furthermore, the findings indicated that speakers delivered jokes more vividly only when interacting with an attentive listener, and when they anticipated the listener to be attentive.

Malisz et al. (2016) also replicated the distraction task with German speakers, telling two holidays stories. The feedback function was not annotated into generic and specific categories, but rather in terms of functions, encompassing expressions of *perception*, *understanding*, and *acceptance/agreement*, along with feedback conveying attitudes or introducing/ending new topics or discourse segments. A total of 20 dyads were annotated, and the analysis conducted revealed a general decrease in feedback frequency when listeners were distracted. The study also delved into the prosodic and gestural aspects of feedback. Distracted listeners exhibited feedback with less pitch (F0) variability but higher intensity variability. Additionally, their use of verbal feedback decreased, and they tended to produce feedback in the gestural modality.

2.2. Our Hypothesis

Our work, based on the results from (Bavelas et al., 2000), has been designed for going one step forward in our understanding of feedback.

Starting from the well-established distinction between generic and specific feedback (Stivers, 2008; Tolins and Fox Tree, 2014; Bertrand and Espesser, 2017), we focus on how the listener's attention influences various aspects of the conversation, ranging from feedback production, feedback components, feedback perception, participants engagement and its impact on the speaker's discourse.

In (Bavelas et al., 2000), narrators are told that the listening participants are going to look for something in their speech, but they don't know what. In our study, we refrain from informing the narrators that the other participants are engaged in anything other than listening. Instead, we provide listeners with a distinct instruction, namely, *"the storyteller should not realize it"* (for a detailed procedure, refer to 3.1).

Our hypotheses are outlined as follows: **Hyp. #1**: We expect to find a similar frequency of generic feedback in both distracted and normal conditions, as in free conversation. **Hyp. #2**: We expect a lower frequency of specific feedback from participants in distracted conditions compared to listeners in normal conditions, as well as during the free conversation task. **Hyp. #3**: We expect that the quality of storytelling, as rated by a third party, will be lower in the distracted condition compared to the normal condition. **Hyp. #4**: We hypothesize a positive correlation between the frequency of specific feedback and the quality of storytelling. **Hyp. #5**: We expect that feedback produced by distracted listeners are less elaborate, in terms of form, than the feedback produced by normal listeners.

We support the idea that generic and specific feedback opportunities are based on different feedback-inviting features of the main speaker (Boudin et al., 2021). We also believe that generic feedback is based on lower-level features of the main speaker than specific feedback. As a result, we believe that it will be more difficult for distracted listeners to identify and capture the opportunities for specific feedback than for generic feedback.

Moreover, since the listener must avoid appearing distracted, we can expect them to frequently, even unconsciously, produce generic feedback. Indeed, because generic feedback predominantly manifests through nodding and/or vocalizations, making it appear as a "safer" strategy to maintain the appearance of attentive listening.

However, we do anticipate an impact on the form of feedback. Participants who are cognitively occupied in a distracting task are expected to produce feedback with less variability, reduced richness, and shorter duration. This expectation is rooted in the belief that the less engaged a speaker is during a conversation, the less rich and complex their feedback is likely to be. In the present case, distracted listeners are not fully engaged in the conversation due to their hidden t-counting task.

In this study, we replicate the distraction task introduced by (Bavelas et al., 2000), originally designed to validate the generic/specific distinction and demonstrate the collaborative role of listeners. However, our objective is to explore various aspects of feedback. More precisely, feedback opportunities (i.e. identifying when feedback can occur) (Ward and Tsukahara, 2000; Morency et al., 2010; Ruede et al., 2019), features eliciting feedback (Allwood and Cerrato, 2003; Terrell and Mutlu, 2012; Gravano and Hirschberg, 2011; Ferre and Renaudier, 2017; Brusco et al., 2020), the feedback form, and the listening styles (individual variations in listening strategies and characteristics). These parameters hold significant relevance to enhance human-machine interaction (Gratch et al., 2006; Bevacqua, 2013; Axelsson et al., 2022). Our study provides a detailed description of the conditions and parameters for a natural behavior during conversations, ultimately guiding us towards higher-quality human-machine interactions.

3. The SMYLE Dataset

SMYLE is an audio-video corpus in French including neuro-physiological recordings. As a first step, in this paper, we only analyze audio-video signal data. It contains 16h of recordings, with 30 pairs of participants engaged in dyadic 1) **face-to-face storytelling** (8.2h) followed by 2) a **free conversation** task (7.8h). The storytelling task comprises two conditions: a storyteller talking with a "*normal*" or a "*distracted*" listener.

Participants Sixty participants took part in the experiment, with a mean age of 22.77 years. The group included 43 females and 17 males, consisting mostly of students, from various fields. All participants were native French speakers with no reported neurological or language disorders. Participants received compensation of €30 each. None of the participant dyads knew each other before the experiment. The storytelling task of one dyad was dropped for technical reasons. Currently, 50 participant dyads have been fully annotated (see 4) for the analyses conducted here. Another dyad was dropped due to a lack of understanding of the instructions by the listener, leaving 24 dyads for analysis, 12 in the normal condition and 12 in the distracted condition. The normal condition is composed of 8 female-female dyads, 2 female-male dyads and 2 male-male dyads. The distracted condition is comprises 7 female-female dyads, 4 female-male dyads and 1 male-male dyad.

Experimental Set-up and Equipment Participants were placed in an anechoic chamber, seated face-to-face 130 centimeters apart. High-quality cameras positioned in front of each participant recorded their interactions, and each participant wore a headset microphone for clear audio capture (the set-up is illustrated Figure 1). The electrophysiological signal was recorded using Biosemi 64 active electrodes, and physiological data were collected with Empatica E4 wristbands. A green background and spotlights were used for optimal video analysis. The EEG systems and physiological sensors were synchronized for data collection (for a more detailed description see (Boudin et al., 2023a).

Figure 1: Picture of the set-up with Participant A (listener) on the left and Participant B (main speaker) on the right.

3.1. Tasks

Each dyad participates to two parts: one experimental task based on a storytelling activity (with two conditions: control and distracted), followed by a free conversation.

Storytelling task The storytelling task comprises a control condition and a distracted condition. Each participant is randomly assigned a specific discursive role for the entire task, either storyteller or listener. To ensure an adequate duration of interactions, storytellers were instructed to narrate three stories to the other participant: #1 retelling the pear story video (Watson-Gegeo, 1981), #2 narrate the pitch of a movie/book/video game and #3 relate their favorite vacation. For a detailed description of the instructions, please refer to the appendices. Storytellers were asked to tell the stories one after the other. In contrast to (Bavelas et al., 2000) but in line with the results of (Kuhlen and Brennan, 2010), the storytellers were only informed that the other participant would have to summarize the stories after the experiment, but were not informed of any additional tasks.
Listeners in the normal condition were informed that their experimental partner would tell three stories. Participants were instructed to listen carefully and to freely react, speak, ask questions during the storytelling, and to quickly summarize the stories at the end of the experiment. For the distracted condition, supplementary instructions were given to the listeners, asking them also to count all words produced by the storyteller that start with the sound /*t*/, and to press a pressure plate with their foot. Unlike (Bavelas et al., 2000), we decided to use the foot rather than the hand in order to preserve hand movements during the conversations. The said plate was actually a trick to encourage them to do the task well. Finally, we told them that the storyteller should not discover this hidden task.

At the end of the storytelling task, each participant is requested to complete a questionnaire tailored to their respective roles (the questionnaires are included in the appendices). This questionnaire is designed to assess participants' level of engagement, as well as the perceived engagement of other participants. It also includes an assessment of narrative quality and listening style. Distracted listeners are asked to indicate the total number of "t-words" counted.

Free conversation In the second part of the experiment, participants were instructed to engage in a 15-minute **free conversation**. They were asked to initiate the conversation with a debriefing of the first task. This debriefing step served multiple purposes: allowing distracted listeners in the distracted condition to reveal their distraction task to the storyteller, facilitating a smooth transition to the free conversation topic, and gathering valuable feedback from participants (their overall experience, task success, encountered difficulties, etc.).

3.2. Procedure

For installation, participants were assigned to separate rooms and provided informed consent. They were informed that the study aimed to investigate spontaneous conversations between strangers. Participants received written and oral instructions before EEG equipment setup. The storyteller watched the pear story video while EEG cap placement occurred. Both participants then moved to an anechoic chamber without prior interaction. After camera and microphone adjustments, EEG signals were verified. Participants initiated the first task upon the beacon signal and concluded it with the same signal. A brief break followed, along with the online questionnaire. The second task began and ended also with the beacon signal, resulting in a total experiment duration of approximately 2 hours.

4. Annotations

We performed automatic and manual annotations on the corpus. Manual annotations and correction have been made by 3 experts annotators.The annotation procedure is schematically illustrated in Figure 5.

Storytelling Quality Bavelas et al. (2000) primarily concentrates on analyzing the quality of story endings in the context of near-miss incident, we establish a method for evaluating narrative quality across the entire story. Three narratives have been chosen, each presenting implications and denouement of varying salience, which will be examined in future work.

The annotators rated the following 6 criteria using a scoring system ranging from 0 to 3. **Level of detail:** From 0, lacking or excessively detailed to 3, perfectly balanced with relevant details. Inter-rater agreement (κ) = 0.27. **Clearness:** From 0, unclear or disjointed to 3, perfectly clear (κ) = 0.33. **Story ending:** From 0, abrupt or never-ending to 3, perfect ending (κ) = 0.32. **Rhythm:** From 0, too slow or too fast to 3, consistent pace throughout the story (κ) = 0.15. **Interest in the story:** From 0, uninteresting to 3, very interesting (κ) = 0.23. **Comfort of the speaker:** From 0, not comfortable at all to 3, very comfortable (κ) = 0.60.

Given the inherent subjectivity of such annotations, all storytellers were assessed by the raters, we next compute the storytelling score by averaging all scores. The annotators were blind to the condition and instructed to evaluate each criterion based on a precise definition, after watching the video only once. This was their initial annotation task before becoming familiar with the corpus. The annotations were conducted using the audio and video of the storyteller.

Speech and acoustic annotations As an initial step, speech have been automatically transcribed into Inter-Pausal-Unit (IPU)[1] (see Boudin et al. (2023a)). The annotators then manually corrected both IPU segmentation and transcription incorporating additional details such as laughter, laughing pronunciation, repetitions, disfluences, broken words, specific pronunciation and elision (Blache et al., 2017).

Phonemes, tokens and syllables aligned to the audio signal, were automatically extracted using the SPPAS software (Bigi, 2012, 2015). Part-of-Speech were automatically annotated using the Marsatag software (Rauzy et al., 2014; Amoyal et al., 2022). An example is given in the appendix, figure 6.

Mimo-gestural annotations Gaze and smile have been automatically annotated using the HMAD software (Rauzy and Goujon, 2018; Rauzy and Amoyal, 2020, 2022). Currently, only gaze have been manually corrected. Head movements have been manually encoded. We used the following labels: **Gaze**: *look at the other participant, do not look at the other participant*. **Smile**: *Neutral face, Low Intensity Smile, High Intensity Smile*. **Head**: *nod, shake, tilt, other*.

Feedback annotations Feedback have been manually annotated by one of the authors into generic and specific feedback type. We define feedback as any response produced by one speaker in reaction to the production of the other speaker, except when providing answers to explicit questions. These reactions can be vocal, verbal, or mimo-gestural. The type tagging is based on the definition between generic and specific feedback of (Bavelas et al., 2000) as describe in 2.1.

In each feedback instance, we perform an accurate annotations of the feedback components. Our annotations encompass various aspects, including the movements of the head, wince, movements of the eyebrows, shoulder shrug, and gestures made with the hands. Concerning lexicalized feedback, we specify the type of lexical content used within the feedback. We used the following labels: **Hands:** *deictic, metaphoric, beat, iconic*. **Eyebrows:** *raised, frowned*. **Shoulders:** *shrug*. **Wince:** *pout, wrinkle nose, frown eyes, big eyes, roll eyes, bite the lips, stretched mouth, open jaw*. **Lexical type:** *interjection, prototypical form, completion, repetition, reported speech, clarification request, information question, the listener does not know*. An example is given in the appendix, figure 7.

Annotation of conversational roles Annotation of conversational roles is herein achieved by assigning to the two participants the conversational role they play in the interaction, i.e. each participant is whether in a "Speaker" role leading the conversation or in a "Listener" role listening to the main speaker and possibly producing feedback. As long as the feedback annotation have been properly carried out, a deterministic heuristic can be used to infer the conversational roles. The algorithm is fed in input with the speech activity and feedback production of both participants and provides in output a 4 levels annotation describing all along the interaction the combination of the individual conversational roles. Label q_{SL} stands for participant 1 in the role of "Speaker" and participant 2 in the role of "Listener", label q_{LS} draws the reverse situation, label q_{SS} occurs when both participants are simultaneously "Speaker" (e.g. speech overlap areas) and q_{LL} when both are listening. We applied our automatic tool on the totality of SMYLE interactions

[1] defined as speech segments separated by a silent pause of at least 200ms.

with the speech activity provided by SPPAS (Bigi, 2012, 2015) and the feedback productions manually annotated as explained above.

5. Data Description

In this section, we present descriptive statistics for the 48 fully annotated speakers (12.75h, 6.62h from the storytelling task and 6.13h from the free conversations), categorized by conversational role (storyteller or listener) and by condition. As an initial exploration of the condition's impact, we have conducted t-tests between each discursive role to compare the two conditions during the storytelling task. Tables showing the number of items per annotation, mean duration, standard deviation of duration and observed frequency by condition and role for the storytelling task and for the whole corpus are provided in the appendix.

We hypothesize that the distracted condition may result in less "rich" production from both the speaker and the listener, primarily due to the disrupted listening quality of the listener. This disruption could lead to shorter item duration and decreased item frequencies.

When comparing the frequency and the mean duration per participant of each type of item between storyteller in normal condition and storyteller in distracted condition, we found a significant effect only for nod duration. The independent samples t-test revealed a significant difference in test scores between the nod duration of storytellers in **normal condition** (M = 0.94, SD = 0.21) and **distracted condition** (M = 0.73, SD = 0.14), t(18.89) = -2.90 ; p = 0.00928 ; 95% confidence interval = [-0.36 -0.06]. The effect size, as measured by Cohen's d, was d = -1.18, indicating a large effect. Storytellers in distracted condition produces shorter nods than storytellers in normal condition.

When comparing the frequency and the mean duration per participant of each type of item between listeners in normal condition and storyteller in distracted condition, we found a significant effect also for nod duration. The independent samples t-test revealed a significant difference in test scores between the nod duration in **normal condition** (M = 1.23, SD = 0.35) and **distracted condition** (M = 0.94, SD = 0.26), t(20.31) = -2.38 ; p = 0.027 ; 95% confidence interval = [-0.55 -0.04], the effect size, as measured by Cohen's d, was d = -0.97, indicating a large effect. Listeners in distracted condition produces shorter nods than listeners in normal condition. We also found that **distracted listeners** laugh less frequently (M = 1.05 per minute, SD = 0.75) than **normal listeners** (M = 1.92 per minute, SD = 1.16), t(18.87) = -2.17 ; p = 0.0429 ; 95% confidence interval = [-1.70 -0.03]. The effect size, as measured by Cohen's d, was d = -0.88, indicating a large effect.

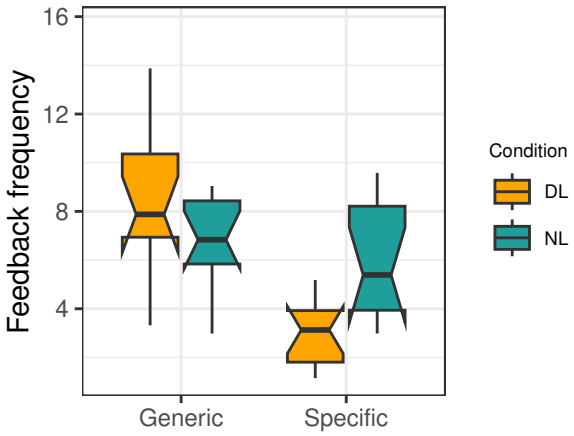

Figure 2: The mean frequency of generic and specific feedback by condition: Distracted Listener (DL), Normal Listener (NL) during the storytelling task.

6. Results

6.1. Feedback Frequency

We first compared the feedback frequency of generic and specific feedback between attentive listeners and distracted listeners during the storytelling task. One original aspect of the SMYLE corpus is that, unlike previous works in this domain, the free conversation task offers the opportunity to compare the distracted listener with their listening position without distraction during free conversation.

We therefore calculated the generic and specific feedback frequency per minute during storytelling i.e. frequency $=$ total number of feedback$/$task duration (min), illustrated in figure 2. During free conversations, the conversational roles assumed by participants change rapidly and very frequently. To obtain the feedback frequency for each participant, it is important to calculate it during the time when they are actually serving as the listener. We therefore used the conversational role annotations when *the listener of the storytelling task* is in the listener position (labels q_{SL} and q_{LL}) during free conversation as the corrected total time of free conversation. Consequently, the feedback frequency in free conversation is calculated as frequency $=$ total number of feedback$/$duration$(\text{q}_{\text{SL}} + \text{q}_{\text{LL}})$.

As indicated in hypothesis 1 and 2, we expect to find an equal frequency of generic feedback in both conditions and a higher frequency of specific feedback in the normal condition compared to the distracted condition. We expect the same effects when we compare distracted listeners during narration and free conversation.

The independent samples t-test revealed no significant effect for **generic** feedback frequency be-

tween **normal** and **distracted** conditions during the **storytelling task** (p = 0.36), indicating that a similar frequency of generic feedback is observed between the two conditions. The independent samples t-test revealed a significant difference in test scores between the **specific** feedback frequency in **normal condition** (M = 4.24, SD = 1.51) and **distracted condition** (M = 2.94, SD = 1.4), t(21.88) = -2.18 ; p = 0.04014 ; 95% confidence interval = [-2.53 -0.06] during the **storytelling task**. The effect size, as measured by Cohen's d, was d = -0.89, indicating a large effect.

The independent samples t-test revealed a significant difference in test scores between the frequency of **specific** feedback during the **storytelling task** among participants in the distracted condition (M = 2.94, SD = 1.4) and their frequency during the **free conversation** task (M = 5.91, SD = 2.37), t(17.87) = -3.74 ; p = 0.0015 ; 95% confidence interval = [-4.65 -1.30]. The effect size, as measured by Cohen's d, was d = -1.52, indicating a large effect. The independent samples t-test revealed no significant effect for participants in **normal** condition between the **storytelling task** and the **free conversation** (p = 0.11). These results indicate a reduced frequency of specific feedback when the listener is distracted.

To summarize, we did not find significant differences for generic feedback, but we found a significant effect for specific feedback between distracted listeners and attentive listeners during the storytelling task. Additionally, we found a significant effect between the storytelling task and the free conversation for distracted listeners, but not for attentive listeners. Hypotheses 1 and 2 have thus been confirmed. Results are summarize in table 3.

6.2. Feedback Components

In this section, we present the analysis of feedback components in both normal and distracted listening conditions during the storytelling task. We define feedback components as individual elements annotated within the feedback interval, as outlined in the feedback annotation section 4.[2]

[2]For example, if a feedback specific is produced with a *completion*, an *interjection*, a *shake* and *raised eyebrows*, the feedback is composed of 4 components.

Cond	Gen freq.	Gen dur.	Spe freq.	Spe dur.
DC ST	8.76 $\pm$ 3.43	0.90 $\pm$ 0.40	2.94 $\pm$ 1.40	1.79 $\pm$ 0.64
NC ST	7.69 $\pm$ 1.97	1.22 $\pm$ 0.35	4.24 $\pm$ 1.51	2.00 $\pm$ 0.51
DC FC	7.17 $\pm$ 3.13	0.88 $\pm$ 0.31	5.91 $\pm$ 2.02	2.01 $\pm$ 0.43
NC FC	7.31 $\pm$ 3.13	1.19 $\pm$ 0.34	5.44 $\pm$ 2.02	1.89 $\pm$ 0.20

Table 1: *Feedback frequency and duration for generic and specific type, for normal condition (NC) and distracted condition (DC) and during the storytelling (ST) and the free conversation (FC).*

Our fifth hypothesis is that feedback produced during the distracted condition will be composed of a minimal quantity of components (less rich or less complex) compared to feedback produced during the normal condition, for both generic and specific feedback.

We tested whether the component quantity for generic and specific feedback follow a parametric distribution in both condition using the Shapiro test, all produced a p-value < 2.2e-16. These results shows the non-normal distribution of our data.

We therefore conducted the non-parametric Wilcoxon-Mann-Whitney test to compare the components quantity of **generic** feedback between the **normal** (M = 1.43, sd = 0.61) and **distracted** (M = 1.34, sd = 0.57) listening conditions during the storytelling task. The results revealed a significant difference between the two conditions (W = 1430736 and p-value = 1.227e-05).

Similarly, a Wilcoxon-Mann-Whitney test was performed to compare the components quantity of **specific** feedback between **normal** (M = 3.01, sd = 1.63) and **distracted** (M = 2.99, sd = 1.69) conditions during the storytelling task. Surprisingly, no significant difference was observed, (p-value of 0.7271).

In other words, generic feedback from distracted listeners is significantly produced with less components than from normal listeners but there is no significant differences for specific feedback.

Next, we computed all feedback components combinations observed during the storytelling task for both feedback type. We dropped feedback containing only smile for this analysis because we do not have a manual correction yet.

We had **1,425 generic** feedback in **normal** condition realized with **52 unique component combinations** and **1,851 generic** feedback in **distracted** condition with **45 unique combinations**. We had **814 specific** feedback in **normal** condition with **350 unique component combinations** and **698 specific** feedback in **distracted** condition with **339 unique combinations**.

To compare the most frequently used combinations between the two conditions, we present in Figure 3 the graphical representations of components combination. It displays 95% of the generic feedback combinations for the normal and distracted conditions and 30% of the specific feedback combinations for the normal and distracted conditions.

We observe that for generic feedback in the distracted condition, only 3 combinations are sufficient to represent 95% of the feedback produced, namely *nods*, *continuers*, and *nods combined with continuers*. Comparing this with the normal con-

dition, the first 3 combinations are the same, but we also find additional combinations. These additional combinations are intriguing because they represent generic feedback with a slight specific connotation (some components may render the feedback type ambiguous, in which case we annotate in generic if the component's intensity is low). Among them are shakes, prototypical forms (e.g., "*ah oui*"), raised eyebrow(s), and laughter. As noted by (Bavelas et al., 2000), laughter can be used as both specific and "appreciative generic responses" or to express politeness. The presence of shakes is more surprising, but during the annotations, we observed that they could occur when the main speaker is shaking their head just before or during the feedback. This could be a form of nod that is altered into a shake through alignment or mimicry mechanisms (Pickering and Garrod, 2013).

Turning to combinations for specific feedback, laughter is the most common response in both conditions. It's not surprising that laughter is produced despite the distraction, as if the main speaker laughs, it is something quite noticeable and salient in the conversation, making it highly communicative even during distraction. This aspect deserves further exploration.

For the remaining feedback combinations, while there are slightly fewer of them in the distracted condition compared to the normal condition, we do not observe any significant differences in the combinations. Further in-depth analyses are warranted, such as the integration of acoustic intensity in lexicalized feedback or gestural intensity.

It appears that the complexity and composition of feedback affect only generic feedback and not specific feedback. Therefore, Hypothesis 5, stipulating that "feedback produced by distracted listeners will be less elaborate than the feedback produced by normal listeners", is only validated for generic feedback.

6.3. Storytelling Quality Assessment

We present in (Boudin et al., 2023a) the scores of the storytelling quality from the third-party annotations for all the storytellers. We present here the results of the storytelling quality for the 24 storytellers used for this analysis. We test in this section Hypothesis 3, that the quality of storytelling, as rated by a third party (see 4), will be lower in the distracted condition compared to the normal condition.

The independent samples t-test revealed a significant difference in test scores between storytellers in **normal condition** (M = 2.47, SD = 0.36) and storytellers in **distracted condition** (M = 1.58, SD = 0.8), t(15.38) = -3.52 ; p = 0.002968 ; 95% confidence interval = [-1.42 -0.35]. The effect size, as measured by Cohen's d, was d = -1.44, indicating a large effect. This means that storytellers in a distracted condition are less good at telling stories. Figure 4 depictes the mean scores of storytelling quality per criterion in the normal condition and distracted condition.

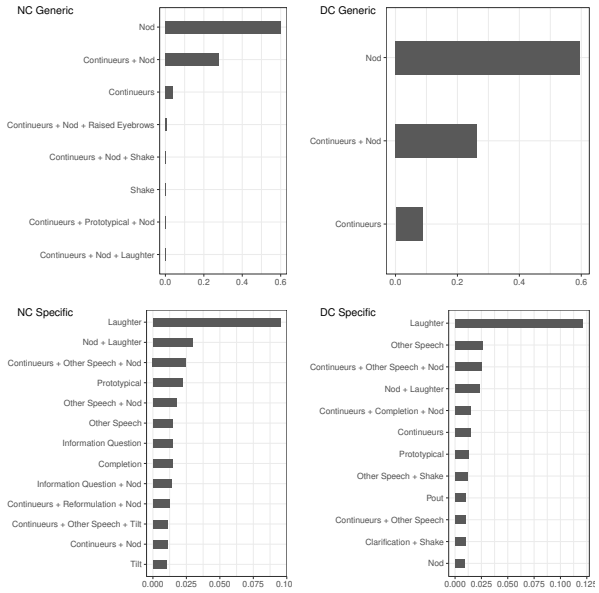

Figure 3: Graphical representation of the most common feedback combination for 95% of generic feedback and 30% of specific feedback according to the listening condition, i.e. normal (NC) or distracted (DC).

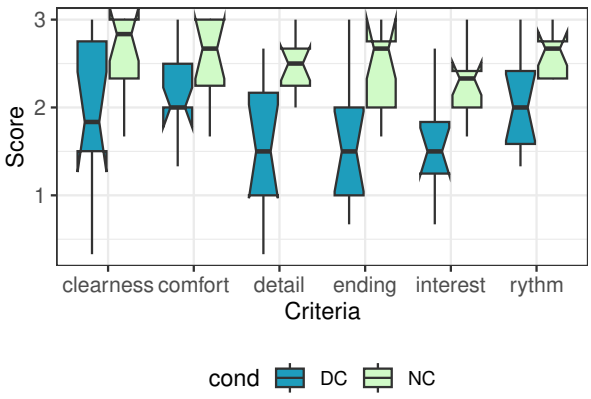

Figure 4: Mean score of storytelling quality per criterion in the normal condition (NC) and distracted condition (DC).

6.4. Correlation Between Feedback Frequency and Storytelling Quality

In order to see the impact of the feedback frequency on the quality of the storytelling (hypoth-

esis 4), we look at the correlation between the storytelling quality and the feedback frequency for generic and specific feedback and by condition.

Following the assertions made by (Bavelas et al., 2000), the quality of the storytelling is influenced by the behavior of the listener, who acts as a "co-narrator" by producing specific feedback in particular. However, on the preliminary analysis conducted in (Boudin et al., 2023a) on SMYLE, we did not observe a correlation between narration quality and the frequency of specific feedback.

One would typically expect the narrator to be affected when facing a distracted listener, which should in turn impact the quality of his/her discourse. The correlation test did not yield any significant relationship between storytelling quality and the frequency of generic feedback (normal condition: ρ = -0.03, CI = [-0.34, 0,3], distracted condition: ρ = 0.04, CI = [-0.29, 0,35]). Nevertheless, in **normal condition**, we identified a positive correlation between the frequency of specific feedback and narration quality (ρ = 0.57, CI = [0.3, 0,75]). Conversely, in **distracted condition**, we found a negative correlation (ρ = -0.49, CI = [-0.7, -0,2]) between the frequency of specific feedback and storytelling quality.

We observe in this study and in (Boudin et al., 2023a) that the less specific feedback listeners provide, the lower the quality of the storytelling. In line with these results, the negative correlation suggests that specific feedback is being produced inappropriately in the distracted condition, either in terms of timing, content, or form. Nonetheless, we cannot definitively conclude on these interpretations without an assessment of the appropriateness (i.e. congruence) of the provided feedback. To gain a better understanding of the listener's intent and the impact of feedback on the narration, it is essential to evaluate to what extent the feedback aligns with the ongoing conversation and the specific context within the distracted condition. Assessing congruence can provide valuable insights into the listener's responses and their role in the overall interaction.

Another possibility is that, in response to realizing that the interaction is not proceeding smoothly due to distraction, the listener may make an effort to compensate by producing more specific feedback. In this scenario, the listener might believe that offering specific feedback can help repair the disrupted interaction, even though this increased specificity doesn't always lead to improved storytelling quality. This hypothesis suggests that the listener's intention is to enhance the interaction, but the outcome may not align with their expectations.

7. Conclusion

In the quest to advance our comprehension of the role and quality of feedback in conversations, we present a unique dataset that replicates the distraction task introduced in the original study by (Bavelas et al., 2000). Our objective in this study is to investigate how the perturbation of the listener's perception impacts feedback frequency, feedback form and storytelling quality.

Our findings reveal that distraction affects both types of feedback, but in distinct ways. Generic feedback becomes less elaborate when the listener is distracted, yet its frequency remains comparable to that in a normal conversation or during free conversation. On the other hand, specific feedback shows no alteration in the complexity of its form but becomes less frequent in distracted conditions. As in (Bavelas et al., 2000), we observe a decline in the quality of storytelling when the listener is distracted (Boudin et al., 2023a). Moreover, we found a positive correlation between specific feedback frequency in normal condition and a negative correlation when the listener is distracted.

These results suggest three important directions for further research. Firstly, to study the different characteristics of the main speakers eliciting generic and specific feedback using predictive models of feedback. Secondly, to investigate the cognitive mechanisms of feedback production according to the feedback type and form, in particular using EEG signal. Thirdly, to further investigate how a distracted listener influences the speech of the main speaker.

Overall, SMYLE corpus represents a valuable resource for various applications. It can be used for measuring conversational engagement and developing strategies for attentive listening robots. In addition, it offers a unique opportunity to study conversational styles, listening styles and alignment, a topic at the forefront of the research field of interactional linguistics.

Acknowledgements

This work, carried out within the Institute of Convergence ILCB, was supported by grants from France 2030 (ANR-16-CONV-0002) and the CNRS MITI, the Cognition Carnot Institute and the ANR (Project COPAINS—ANR-18-CE33-0012). This dataset has been recorded in the soundproof room of the CEP experimental platform (LPL, AMU-CNRS, Aix-en-Provence, France). We would like to thank the team of annotators.

8. Bibliographical References

Jens Allwood and Loredana Cerrato. 2003. A study of gestural feedback expressions. In *First nordic symposium on multimodal communication*, pages 7–22.

Mary Amoyal, Roxane Bertrand, Brigitte Bigi, Auriane Boudin, Christine Meunier, Berthille Pallaud, Béatrice Priego-Valverde, Stéphane Rauzy, and Marion Tellier. 2022. Principes et outils pour l'annotation des corpus. *Travaux Interdisciplinaires sur la Parole et le Langage*, 38.

Agnes Axelsson, Hendrik Buschmeier, and Gabriel Skantze. 2022. Modeling feedback in interaction with conversational agents—a review. *Frontiers in Computer Science*, 4.

Janet B Bavelas, Linda Coates, and Trudy Johnson. 2000. Listeners as co-narrators. *Journal of personality and social psychology*, 79(6):941.

Roxane Bertrand and Robert Espesser. 2017. Co-narration in french conversation storytelling: A quantitative insight. *Journal of Pragmatics*, 111:33–53.

Elisabetta Bevacqua. 2013. A Survey of Listener Behavior and Listener Models for Embodied Conversational Agents. In Matej Rojc and Nick Campbell, editors, *Coverbal Synchrony in Human-Machine Interaction*, pages 243–268. CRC Press Inc.

Brigitte Bigi. 2012. SPPAS: a tool for the phonetic segmentation of speech. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 1748–1755, Istanbul, Turkey. European Language Resources Association (ELRA).

Brigitte Bigi. 2015. Sppas-multi-lingual approaches to the automatic annotation of speech. *The Phonetician. Journal of the International Society of Phonetic Sciences*, 111(ISSN: 0741-6164):54–69.

Philippe Blache, Roxane Bertrand, Gaëlle Ferré, Berthille Pallaud, Laurent Prévot, and Stéphane Rauzy. 2017. *The Corpus of Interactional Data: A Large Multimodal Annotated Resource*, pages 1323–1356. Springer Netherlands, Dordrecht.

Auriane Boudin, Roxane Bertrand, Stéphane Rauzy, Matthis Houlès, Thierry Legou, Magalie Ochs, and Philippe Blache. 2023a. Smyle: A new multimodal resource of talk-in-interaction including neuro-physiological signal. In *Companion Publication of the 25th International Conference on Multimodal Interaction*, ICMI '23 Companion, page 344–352, New York, NY, USA. Association for Computing Machinery.

Auriane Boudin, Roxane Bertrand, Stéphane Rauzy, Magalie Ochs, and Philippe Blache. 2021. A multimodal model for predicting conversational feedbacks. In *Text, Speech, and Dialogue*, pages 537–549, Cham. Springer International Publishing.

Auriane Boudin, Roxane Bertrand Bertrand, Stéphane Rauzy, Thierry Legou, Magalie Ochs, and Philippe Blache. 2023b. Smyle. ORTOLANG (Open Resources and TOols for LANGuage) –www.ortolang.fr.

Pablo Brusco, Jazmín Vidal, Štefan Beňuš, and Agustín Gravano. 2020. A cross-linguistic analysis of the temporal dynamics of turn-taking cues using machine learning as a descriptive tool. *Speech Communication*, 125:24–40.

Gaelle Ferre and Suzanne Renaudier. 2017. Unimodal and bimodal backchannels in conversational english. In *Proceedings of the 21st Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*, Saarbrücken, Germany. SEMDIAL.

Nadine Glas and Catherine Pelachaud. 2015. Definitions of engagement in human-agent interaction. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 944–949.

Jonathan Gratch, Anna Okhmatovskaia, Francois Lamothe, Stacy Marsella, Mathieu Morales, R. J. van der Werf, and Louis-Philippe Morency. 2006. Virtual rapport. In *Intelligent Virtual Agents: 6th International Conference, IVA 2006, Marina Del Rey, CA, USA, August 21-23, 2006. Proceedings 6*, IVA'06, page 14–27, Berlin, Heidelberg. Springer-Verlag.

Agustín Gravano and Julia Hirschberg. 2011. Turn-taking cues in task-oriented dialogue. *Comput. Speech Lang.*, 25(3):601–634.

Anna K. Kuhlen and Susan E. Brennan. 2010. Anticipating distracted addressees: How speakers' expectations and addressees' feedback influence storytelling. *Discourse Processes*, 47(7):567–587.

Zofia Malisz, Marcin Włodarczak, Hendrik Buschmeier, Joanna Skubisz, Stefan Kopp, and Petra Wagner. 2016. The alico corpus: analysing the active listener. *Language Resources and Evaluation*, 50:411–442.

Louis-Philippe Morency, Iwan Kok, and Jonathan Gratch. 2010. A probabilistic multimodal approach for predicting listener backchannels. *Autonomous Agents and Multi-Agent Systems*, 20(1):70–84.

Martin J. Pickering and Simon Garrod. 2013. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4):329–347.

Ronald Poppe, Khiet P. Truong, and Dirk Heylen. 2011. Backchannels: Quantity, type and timing matters. In *Intelligent Virtual Agents*, pages 228–239, Berlin, Heidelberg. Springer Berlin Heidelberg.

Stéphane Rauzy and Mary Amoyal. 2020. SMAD: A tool for automatically annotating the smile intensity along a video record. In *HRC2020, 10th Humour Research Conference*, Commerce, Texas, United States.

Stéphane Rauzy and Aurélie Goujon. 2018. Automatic annotation of facial actions from a video record: The case of eyebrows raising and frowning. In *Workshop on "Affects, Compagnons Artificiels et Interactions", WACAI 2018*, page 7 pages, Porquerolles, France. Magalie Ochs.

Stéphane Rauzy, Grégoire Montcheuil, and Philippe Blache. 2014. MarsaTag, a tagger for French written texts and speech transcriptions. In *Second Asian Pacific Corpus linguistics Conference*, pages 220–220, Hong Kong, China.

Stéphane Rauzy and Mary Amoyal. 2022. Automatic tool to annotate smile intensities in conversational face-to-face interactions. *Gesture*, 21(2-3):320–364.

Robin Ruede, Markus Müller, Sebastian Stüker, and lex Waibel. 2019. *Yeah, Right, Uh-Huh: A Deep Learning Backchannel Predictor*, pages 247–258. Springer International Publishing, Cham.

Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.

Tanya Stivers. 2008. Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction*, 41(1):31–57.

Allison Terrell and Bilge Mutlu. 2012. A regression-based approach to modeling addressee backchannels. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, SIGDIAL '12, page 280–289, USA. Association for Computational Linguistics.

Jackson Tolins and Jean E. Fox Tree. 2014. Addressee backchannels steer narrative development. *Journal of Pragmatics*, 70:152–164.

Khiet Phuong Truong, Ronald Walter Poppe, I.A. de Kok, and Dirk K.J. Heylen. 2011. A multimodal analysis of vocal and visual backchannels in spontaneous dialogs. In *Proceedings of Interspeech 2011*, pages 2973–2976. International Speech Communication Association (ISCA). Eemcs-eprint-20721 ; 12th Annual Conference of the International Speech Communication Association, INTERSPEECH 2011, INTERSPEECH ; Conference date: 28-08-2011 Through 31-08-2011.

Nigel Ward and Wataru Tsukahara. 2000. Prosodic features which cue back-channel responses in english and japanese. *Journal of Pragmatics*, 32(8):1177–1207.

Karen Ann Watson-Gegeo. 1981. Wallace l. chafe (ed.), the pear stories: Cognitive, cultural, and linguistic aspects of narrative production (advances in discourse processes, vol. iii). norwood, n.j.: Ablex, 1980. pp. 323. *Language in Society*, 10(3):451–453.

A. Post-experiment questionnaires

Questions	Response Scale
How would you rate the overall quality of your overall narratives/Story 1/ Story 2/ Story 3?	5 Likert-scale
Do you have any comments on the quality of your narration/speech?	Open
During this experiment, what type of listener was your partner?	not very active, moderately active, very active
Did you find that your partner's responses/reactions were timely?	Yes/no
Did you find that the responses/reactions produced by your partner were rather:	Gestural, verbal, both
Did you find that your partner's responses/reactions were consistent with what you were saying?	Yes/no
Did you feel that your partner supported you while you told the stories?	Yes/No
Did you find that the responses/reactions produced by your partner helped you tell the stories?	Yes/No
Do you think your partner was paying attention?	Yes/No
Did you find that your partner's responses/reactions seemed natural and spontaneous?	Yes/No
Have you paid much attention to your partner's reactions?	Yes/No
How would you rate your involvement in the conversation/Story 1/Story 2/ Story 3?	5 Likert-scale
How would you rate the involvement of your partner in the conversation/Story 1/Story 2/ Story 3?	5 Likert-scale

Table 2: Storytellers questionnaire.

Questions	Response Scale
In everyday life (at work, with friends, family...), what kind of listener are you?	not very active, moderately active, very active
During this experience, what type of listener were you?	not very active, moderately active, very active
During this experiment, what type of listener was your partner?	not very active, moderately active, very active
Do you think the responses/reactions you produced during this experiment were made at a time when your partner needed them?	Yes/no
Do you think the responses/reactions you produced during this experiment were those expected by your partner?	Yes/no
Did you find that your reactions were more :	Gestural, verbal, both
Did you feel you were reacting in the same way as in a natural conversation?	Yes/No
Did you feel you were helping/supporting your partner tell his or her stories?	Yes/No
Were you attentive to your partner's speech?	Yes/No
Did you feel you were being forced to react at certain points in the conversation?	Yes/No
How would you rate the overall quality of your partner's overall narration/Story 1/Story 2/Story 3?	5 Likert-scale
Have you paid much attention to your partner's reactions?	Yes/No
How would you rate your involvement in the conversation/Story 1/Story 2/ Story 3?	5 Likert-scale
How would you rate the involvement of your partner in the conversation/Story 1/Story 2/ Story 3?	5 Likert-scale
Describe in a few points what you retained from your partner's first/second/third story	Open
How would you rate your involvement in the conversation/Story 1/Story 2/ Story 3?	5 Likert-scale
How would you rate the involvement of your partner in the conversation/Story 1/Story 2/ Story 3?	5 Likert-scale

Table 3: Normal and Distracted Listeners questionnaire.

Questions	Response Scale
How many words beginning with the /t/ sound did you count?	Open
Did you find your task difficult?	5 Likert-scale
During this experiment, what type of listener was your partner?	not very active, moderately active, very active
Do you think you have successfully completed the entire experiment?	Yes/no
Has this task prevented you from understanding all the information given by your partner?	Yes/no
Were you able to follow everything your partner said?	Yes/no
Did you feel you were reacting in the same way as in a natural conversation?	Yes/No
Did you feel you were helping/supporting your partner tell his or her stories?	Yes/No
Were you attentive to your partner's speech?	Yes/No
Did you feel you were being forced to react at certain points in the conversation?	Yes/No
How would you rate the overall quality of your partner's overall narration/Story 1/Story 2/Story 3?	5 Likert-scale
Have you paid much attention to your partner's reactions?	Yes/No
How would you rate your involvement in the conversation/Story 1/Story 2/ Story 3?	5 Likert-scale
How would you rate the involvement of your partner in the conversation/Story 1/Story 2/ Story 3?	5 Likert-scale
Describe in a few points what you retained from your partner's first/second/third story	Open
How would you rate your involvement in the conversation/Story 1/Story 2/ Story 3?	5 Likert-scale
How would you rate the involvement of your partner in the conversation/Story 1/Story 2/ Story 3?	5 Likert-scale

Table 4: Additional questions for distracted listeners.

B. Annotations

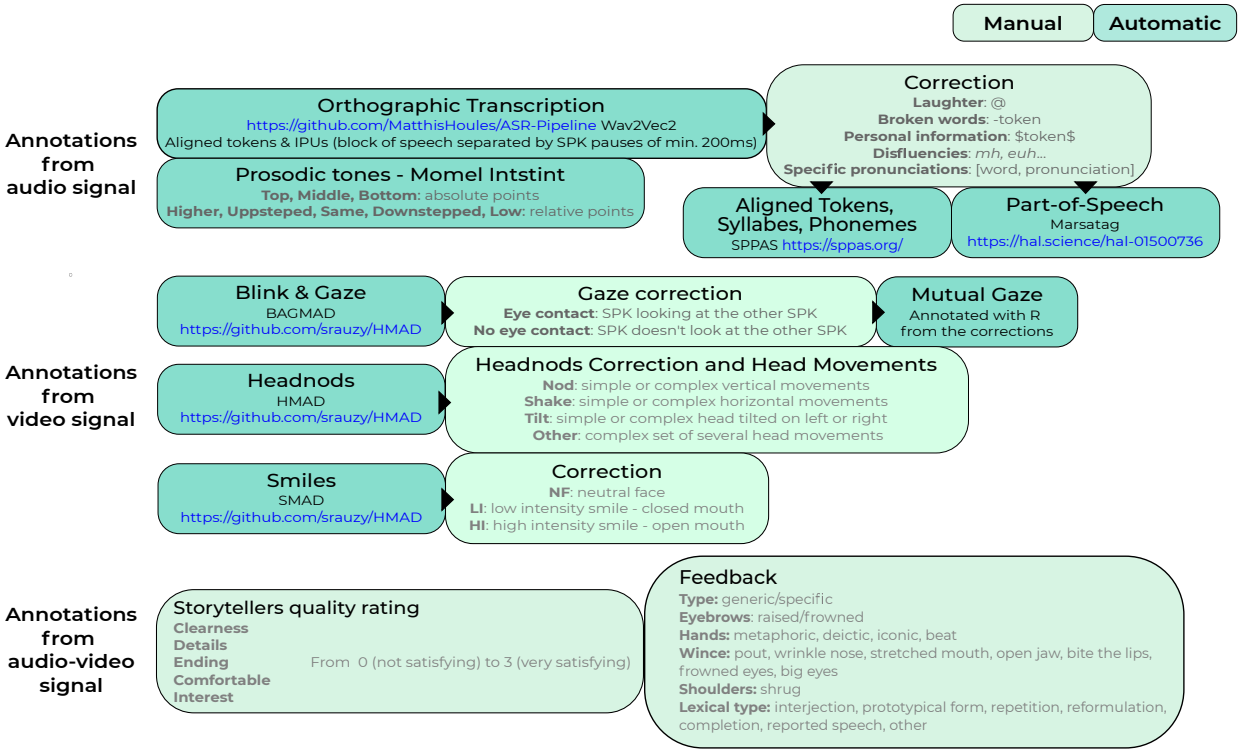

Figure 5: The automatic and manual annotation procedure of SMYLE. Please note that Smiles correction has not been conducted thus far.

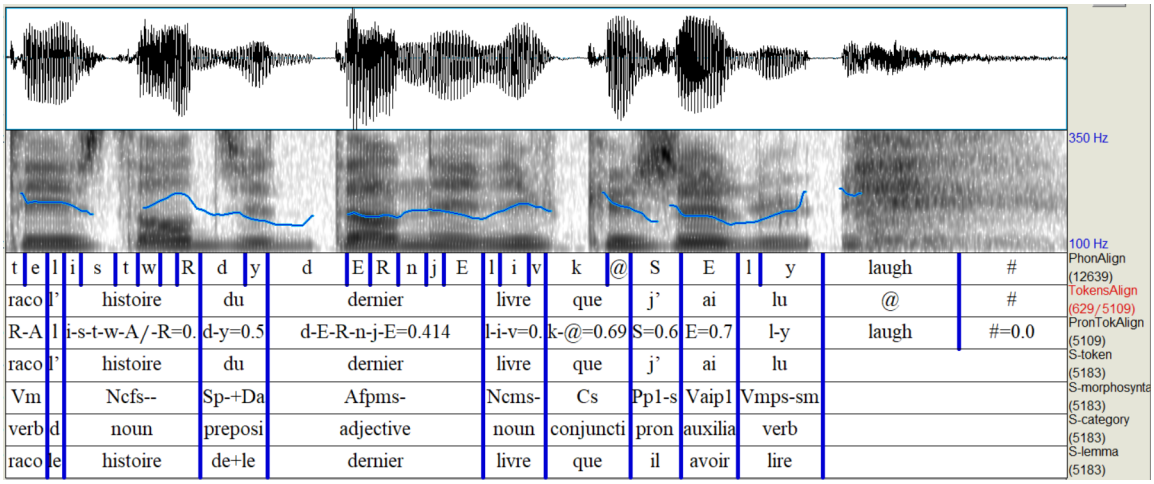

Figure 6: A screenshot of Praat software illustrating the different levels of lexical annotation accomplished through SPPAS and Marsatag, encompassing phonemes, tokens, parts of speech, and lemmas aligned onto the audio signal. The example provided here is "*Je te raconte l'histoire du dernier livre que j'ai lu* (rire)."

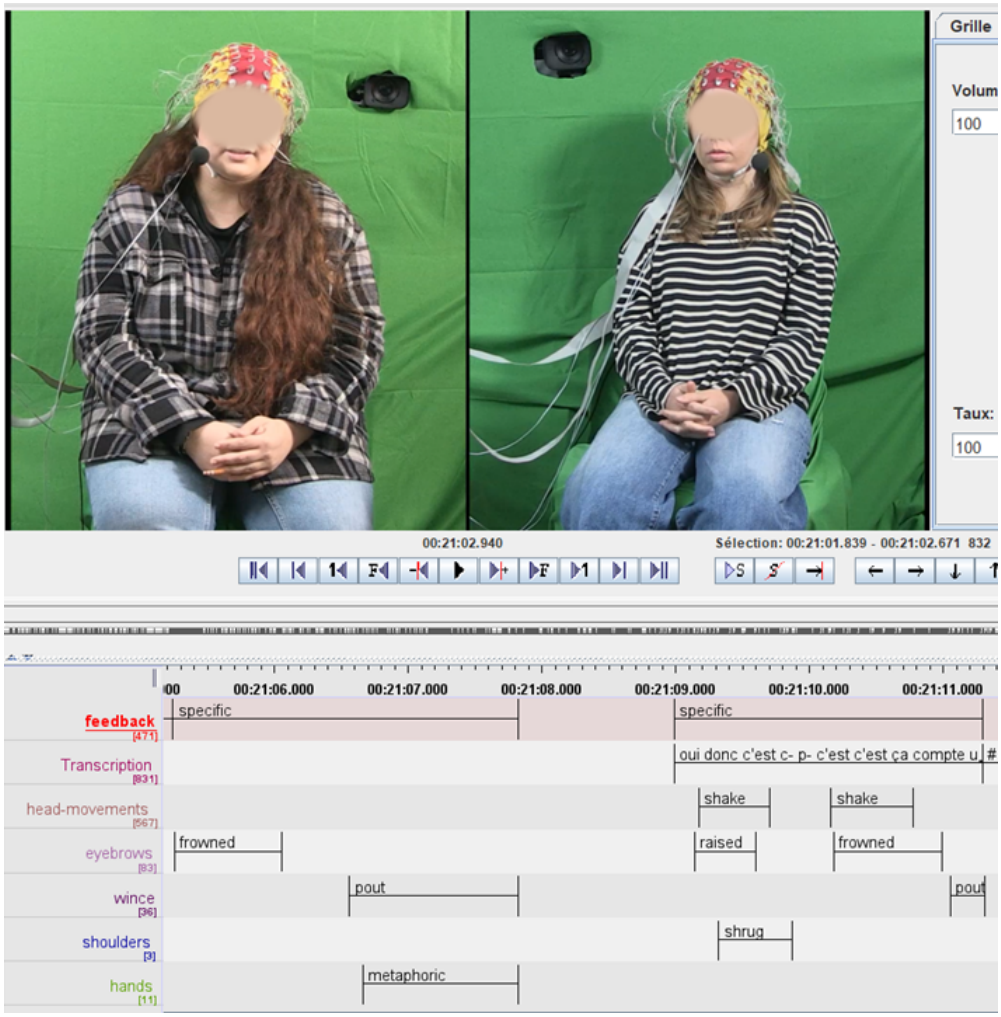

Figure 7: A screenshot of ELAN software illustrating the different levels of annotation of feedback, encompassing transcription, head movements, eyebrow movements, wincing, shoulder shrug, and hand movements.

C. Data Description

Item	Total	Mean dur. (s)	SD dur. (s)	Mean frequency
IPU NC	3,507	2.32	1.89	19.12
IPU DC	4,211	2.21	1.72	19.72
Token NC	33,837	0.24	0.18	184.46
Token DC	40,088	0.23	0.18	187.74
Laughter NC	231	0.55	0.35	1.26
Laughter DC	209	0.47	0.33	0.98
Gaze NC	2,330	3.08	4.34	12.70
Gaze DC	2,658	2.73	3.24	12.45
No Gaze NC	2,317	1.66	1.82	12.63
No Gaze DC	2,656	2.11	2.35	12.44
Nod NC	1,665	0.94	0.70	9.08
Nod DC	1,431	0.75	0.57	6.70
Shake NC	642	1.09	0.71	3.50
Shake DC	1,048	1.04	0.76	4.91
Tilt NC	466	0.87	0.53	2.54
Tilt DC	314	0.62	0.36	1.47
Other NC	149	1.19	0.91	0.81
Other DC	159	1.03	0.59	0.74

Table 5: *The total number of annotated items, the average duration, and standard deviation in seconds, as well as the frequency per minute for participants in the **Storyteller** role during the storytelling task in the **normal** (NC) and **distracted** condition (DC) are presented.*

Item	Total	Mean dur. (s)	SD dur. (s)	Mean frequency
IPU NC	1,573	1.08	1.28	8.58
IPU DC	1,514	0.81	0.86	7.09
Token NC	7,229	0.23	0.16	39.41
Token DC	5,094	0.23	0.18	23.86
Laughter NC	320	0.67	0.57	1.74
Laughter DC	217	0.58	0.47	1.02
Gaze NC	684	14.87	41.3	3.73
Gaze DC	847	13.68	34.78	3.97
No Gaze NC	654	1.15	1.05	3.57
No Gaze DC	834	1.01	1.03	3.91
Nod NC	1,811	1.31	0.96	9.87
Nod DC	2,025	1.05	0.88	9.48
Shake NC	246	1.18	0.78	1.34
Shake DC	222	0.93	0.48	1.04
Tilt NC	261	0.78	0.30	1.42
Tilt DC	140	0.67	0.35	0.66
Other NC	27	1.29	0.85	0.15
Other DC	37	0.80	0.24	0.17

Table 6: *The total number of annotated items, the average duration, and standard deviation in seconds, as well as the frequency per minute for participants in the **Listener** role during the storytelling task in the **normal** (NC) and **distracted** condition (DC) are presented.*

Item	Total	Mean dur. (s)	SD dur. (s)	Frequency
IPU	22,376	1.79	1.67	29.27
Token	173,310	0.23	0.18	226.70
Laughter	2,326	0.63	0.58	3.04
Gaze	13,110	5.26	15.35	17.15
No Gaze	13,023	1.69	1.88	17.04
Nod	12,308	1.02	0.78	16.10
Shake	4,132	1.07	0.76	5.40
Tilt	2,395	0.77	0.45	3.13
Other	645	1.09	0.68	0.84
Generic FB	5,958	1.11	0.84	7.79
Specific FB	3,832	1.99	1.39	5.01

Table 7: The total number of annotated items, the average duration and standard deviation in seconds, and the frequency per minute for the 48 participants in the whole corpus.