Role of prosody on the perception of the " oui " / " yes " feedback in medical context

Mastriani Camilla, Caterina Petrone, Roxane Bertrand, Magalie Ochs

Role of prosody on the perception of the "oui"/"yes" feedback in medical context

Maria Camilla Mastriani [1], *Caterina Petrone*[2], *Roxane Bertrand*[2], *Magalie Ochs*[3]

[1] Università degli Studi di Napoli «Federico II», Napoli, Italy
[2] Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France
[3] LSIS UMR 7296 Aix-Marseille Université, CNRS, ENSAM, Université de Toulon, France
m.mastriani@studenti.unina.it, caterina.petrone@univ-amu.fr, roxane.bertrand@univ-amu.fr, magalie.ochs@lsis.fr

Abstract

This paper focuses on the "oui"/"yes" feedback produced by a patient in the specific medical environment of breaking bad news. In particular, we aim at determining the role of prosody (intonation and temporal delay with which "oui" is produced) in the perception of this feedback. 15 French listeners listened to short human-human interactions between an acting doctor and his patient. They judged the more or less appropriate nature of the "oui" produced by the patient, based on the way it was orally said. The effects of intonation (neutral/ shaken/questioning), delay (short/long), and listener sex (male/female) on judgment scores (1-5) and on the log of reaction times were measured. The results show a role of intonation and delay in a specific part ("phase") of the dialogue (Problem Definition) with female and male listeners being sensitive to different aspects of prosody. This has implications for modeling prosody of a virtual patient's feedbacks in the context of humane-machine interaction.

Index Terms: perception, medical context, feedback, communicative function, intonation, delay, French.

1. Introduction

The perception study presented in this paper is part of a larger project, the Acorformed Project[1], that aims at developing a virtual patient in a virtual reality environment to train doctors to break bad news. Currently, computers are increasingly used in roles that are typically fulfilled by humans, such as virtual tutors in a learning class, virtual actors for social training, or virtual assistants for task realization. When computers are used in these roles they are often embodied by animated cartoon or human like virtual characters, called *Embodied Conversational Agents* (ECA) [1]. This enables a more natural style of communication for the human and allows the computer to avail of both verbal and nonverbal behavior channels of communication.

One of the key elements to create an engaging interaction is the *feedback* behavior of the virtual character. In intelligent virtual agent domain, several researches have demonstrated the importance of embodied conversational agent's feedback for the interaction (e.g. perception of the agent, flow of the conversation, establishment of rapport) [2]. Several computational models have been proposed to automatically predict when an artificial agent should display feedback (as for instance [3, 4]) based on verbal and nonverbal cues of the main speaker. However, few of them have considered the prosody of feedbacks. A virtual agent should be able to adapt the prosody of its feedback depending on the communicative intention it aims at expressing. The development of such an agent requires a better understanding of the role of prosody on the perception of feedback, a particularly important aspect to model patient's verbal behavior in the context of breaking bad news. For this purpose, we have conducted a first experiment in the specific medical context of the project.

Following [5] linguistic feedbacks are mechanisms which enable participants of a conversation to exchange information about four basic communicative functions: contact, perception, understanding and attitudinal reaction. Verbal feedbacks are mostly performed through short utterances such as *ouais* (*yeah)*, *oui* (*yes)*, *mh*, *mhm*, *okay*. Previous studies have investigated the different characteristics of feedbacks for improving the disambiguation of the discourse/pragmatic functions, more particularly at a prosodic level (for a review see [6]). Intonation contours are considered a salient cue to express various types of discourse functions ([7; 8] among others). Among the studies showing the active collaboration from the recipient, several works have specifically shown that the listener's role in storytelling is to provide *appropriate feedback responses* depending on their *generic*/*specific* function (i.e. an understanding/more evaluative function, Bavelas et al 2000; also referred to as c*ontinuers/assessment*, Schegloff , 1982), but also depending on the localization and the timing with which they are produced (Bertrand & Espesser 2017, Bertrand & Priego-Valverde 2017). Similarly, work in the Conversational Analysis framework has shown the importance of the delay with which the interlocutor can provide a typical response. In line with the *preference organization* principle [9]), the interlocutor would tend to express a preferred (expected) response rather than a dispreferred (unexpected) one among different potential alternatives, such as in response to a question or an offer an answer and an acceptation respectively. Different studies have shown that preferred responses would be produced with a smaller delay than dispreferred ones [10; 11]. Yet, little has been done about the prosodic features and the appropriate nature of feedback, and more about their implementation in dialogues involving ECA [12, 13]. Our long-term goal is to test whether the inclusion of prosodic features of feedbacks can improve the interaction with ECA in medical contexts. Before implementing rules in the ECA a first step involving human-human interaction is required. This paper then reports on a preliminary perceptual analysis on the feedback "oui" ("yes") as produced by a human agent playing the role of a patient in short doctor-patient interactions. We wonder to what extent prosodic parameters (intonation and temporal delay) impact the perceptual evaluation of the feedback as the

[1] http://www.lpl-aix.fr/~acorformed/

most appropriate response to a previous context utterance. We expect that "oui" feedbacks with less marked prosody (e.g., produced with neutral/questioning intonation or after a short delay) will be more compatible with contexts allowing the use of "oui" as a simple continuer, while a more marked prosody (e.g., produced with shaking intonation or after a long delay) will be judged as more appropriate after context evoking more emotional engagement. Longer reaction times are supposed to reflect task difficulty and they are expected with inappropriate items. Finally, given that listeners'sex can modulate emotion perception [REF], a preliminary analysis will be reported by splitting the results for male and female listeners..

2. Corpus analysis of feedbacks

Before conducting the main perception experiment, we analyzed a corpus of real training sessions in French medical institutions created within the Acorformed Project (Section 2). Specifically, we made a survey of feedbacks provided by a patient in interaction with a doctor in the context of breaking bad news. Based on these preliminary findings, a perception experiment has been carried out focusing only on the feedback "oui", in which its temporal delay and intonation have been orthogonally manipulated (Section 3). Finally, a general discussion and perspectives are presented (Sections 4 and 5).

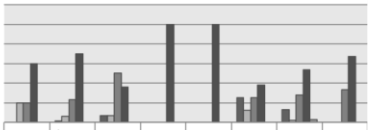
2.1 Corpus description

Six real training sessions collected within the Acorformed Project were analyzed. The traning sessions consisted in dialogues acted between a doctor and a nurse in the role of a patient's relative (for more details about the corpus, see [3]). On average, each dialogue lasted about 15 minutes. Whatever the scenario (variable according to the recording session and the medical institution), each dialogue is structured in five parts or "phases": an *Opening* phase, in which the doctor is welcoming the interlocutor; an *Advert* phase, in which the reason of the meeting is explained; a *Problem Definition* (henceforth, "PD") phase, where doctor provide information about the health conditions of the patient; a *Future Implications* (henceforth, "FI") phase, where possible future complications are described; a *Closing* phase stands for the section of goodbye and leaving (for phases organization in medical contexts: [16]).

2.2 Analysis and results

Following classification in [14], 8 types of feedback have been produced by the patient's relative right after a doctor's utterance. Figure 1 shows their distribution across the five phases. The feedback "oui" is among the most frequent ones (after "ouais"/"yeah") and it is the only one occurring in all phases. Here we will focus on the "oui" feedback in only two phases, PD and FI, where the highest number of "oui" occurs. An example of a sentence produced by the doctor in the PD phase is *"There is an obstacle in his digestive tract*", where the health condition of the patient is described. An example of a sentence produced by the doctor about the possible impacts of the disease and extracted from the IF phase is *"He might not be conscious when he wakes up*". Both sentences are followed by a "oui" feedback which has been spontaneously uttered by the patient's relative.

Furthermore, the temporal delay of the "oui" feedback has been measured in both PD and FI contexts (i.e., the time elapsed before producing the feedback). Such a delay spanned from a minimum of 0 s to a maximum of 0.968 s, with a medium value of 0.391 s chosen as boundary between what we called "short" and "long" delays. The distribution of long and short delays produced by the patient's relative is different between PD and FI phases. Long delays occur in 86% of the cases in IF and in 14% in PD, whereas short delays occur in 47% of the cases in IF and in 53% in PD. during the FI phase, the listener is probably more emotively involved and "spends more energies" in realizing and interiorizing bad information (specific response), that could explain why "oui" occurs after a longer delay. A relatively higher number of short delays in the PD phase compared to the IF phase may be symptomatic of a larger use of "oui" functioning as a continuer feedback (generic response) to show attention to the current discourse and to elicit more information from the speaker.



	ah	d'accord	mm	oh	ok	ouais	oui	voilà
opening		1%	7%			25%	13%	
advert	20%	6%	7%			12%	2%	
problem definition	20%	23%	50%			25%	28%	33%
future implications	60%	70%	36%	100%	100%	38%	54%	67%
closing							3%	

Figure 1: Distribution of feedbacks across the five phases.

3. Perception study

Based on the prior corpus analysis of feedback, the stimuli for the main perception experiment consisted of natural sentences produced by two French native speakers. A rating task was conducted prior to the experiment to select the stimuli.

3.1. Rating for material selection

The stimuli consisted in context-feedback pairs reproducing short doctor-patient's relative interactions. Contexts included 20 natural utterances pronounced by a native French man with a background in linguistics. This speaker played the role of a doctor. The feedback "oui" was produced by a native French woman with a background on prosody. This speaker played the role of the patient's relative. Details about the corpus are presented in section *3.1.1*.

3.1.1. Corpus

Stimuli (context sentences and "oui" feedbacks) were recorded in the anechoic room of the *Laboratoire Parole et Langage* (LPL) in two separate sessions. As for the context sentences produced by the acting doctor, 8 were extracted from the real training corpus and 12 sentences were created from scratch and inspired from the 8 ones extracted from the corpus. Out of the 20 context sentences, half of them are more likely to occur in PD phase while the others in FI phase.

A French female phonetician with a long training on prosody, played the patient's relative receiving bad news from the doctor. She produced the feedback "oui" with three different intonations: *neutral*, *interrogativ*e and *shaken*. To elicit them, she was given explicit instructions: "*you show you are paying attention to the doctor's speech*" for the production of the

"*neutral oui*"; "*you not only show that you pay attention to his speech, but you need clarification or more information to understand the problem*" for the "*questioning oui"*; "*you show not only that you pay attention to his speech produced, but you are shaken*" for the "*shaken oui*".

Specific contexts were given to the speaker to facilitate the elicitation of the different prosodic patterns (e.g., "*the doctor tells you that he himself contacted you on the phone*" for triggering a "*neutral oui*"; *"the doctor tells you he has encountered difficulties during the surgical operation*" for triggering an *"interrogative oui*"; and *"actually, the obstruction of the stomach... it's a cancer*" for triggering a *"shaken oui"*.) In total, we had 3 samples for each intonation of the feedback (i.e., 9 *oui* stimuli), classified "N1", "N2", "N3" (*neutral*), "Q1", "Q2", "Q3" (*questioning*), "S1", "S2", "S3" (*shaken*); we needed only 1 "oui" for any kind (total: 3 *oui* stimuli, the most recognizable in terms of prosody) to build our *corpus*, so we settled a rating task.
As for the temporal delay of the "oui" feedback, employing the 0.0 s bound detected in the Acorformed dialogues as short delay would have been estranging: our corpus is not made up of dialogues but of single "doctor's sentence - delay - patient's relative feedback" interactions. To let the short delays be believable in individual stimuli, for the main experiment, we decided to raise timing to 0.3 s (anyway lower than the medium value of 0.391 s, thus respecting values analyzed in the real dialogues). Timing of long delays has been rounded off to 1s. We recorded by PRAAT two "silences" timed 0.3s and 1s: short delay and long delay.

3.1.2. Participants and procedure

Eight French native speakers, 3 males and 5 females, all students aged between 22 and 25 years, were asked to listen to "oui" samples in isolation (i.e. out of a context). Participants had to identify the prosody of the "oui" choosing between three possibilities (neutral, questioning or shaken). The stimuli were presented through the software PERCEVAL [15].

3.1.3. Results

For "*neutral* oui", N3 was correctly interpreted as *neutral* by 6 listeners. N1 correctly was identified only by 2 listeners and N2 by 4 listeners. As for the *"questioning* oui*"*, Q1 and Q3 have been well identified by all participants, while Q2 was correctly identified by 7 participants. As for the "*shaken* oui", S1 and S2 were correctly identified by 6 listeners over 2, S3 by all of them.
N3, Q3 and S3 were thus employed to build our *corpus* for the main perception experiment.

3.2. Main experiment

Finally, for our *corpus* we had: 20 sets of context utterances produced by a French native speaker in the role of a doctor giving bad news, equally divided between PD and FI contexts, copied from the Acorformed *corpus* or created on its model; 3 sets of "oui" feedbacks differentiated in prosody ,which have been selected after the rating task; 2 delays for each of the 3 "oui" feedback (short = 0.3s and long = 1s).
We combined, respectively, each set of context utterances with both delays and all three intonations of oui: the *corpus* is thus finally composed of 20x6 = 120 stimuli.

3.2.1. Participants and procedure

15 French native speakers, 5 males and 10 females, aged between 18 and 30 years, participated in the main perception experiment. They were asked to listen to the 120 stimuli presented on a laptop through the software PERCEVAL. They were introduced to the experiment by instructions: "*You are going to listen to fragments of conversations between a doctor and a patient's relative"*. The task was to judge the appropriate nature of the feedback "oui" on a scale from 1 ("not at all appropriate") to 5 ("absolutely appropriate") based on the way it was orally said. The judgment was to be expressed through the aid of a button-box provided with 5 buttons.

3.2.2. Results

A series of linear mixed effects models with maximal random structure was separately run for the two phases, "Problem Definition" and "Future Implications", in which we tested the effects of the fixed factors DELAY (short/long), INTONATION (neutral/ shaken/questioning) and LISTENER SEX (male/female) on judgment scores (1-5) and on the log of reaction times. For INTONATION, dummy contrasts were computed with "neutral" as the reference level. LISTENERS and ITEMS were included as random intercepts, with random slopes for each fixed factor. Only significant effects are reported below.

For the phase "Problem Definition", a significant interaction between INTONATION and LISTENER SEX was found on the judgment scores. Specifically, female listeners rated the shaken oui lower than the neutral one [β = -0.97, SE = 0.37, t = -2.5, p = .01] while there was no difference between neutral and questioning oui. On the other hand, male listeners rated the questioning oui lower than the neutral one [β = -1.004, SE = 0.40, t = -2.45, p = .02]; no differences were found between neutral and shaken oui. The factor DELAY and its interactions were not significant. Figure 2 (top) shows the mean judgment scores by prosody, split by listener sex (data are collapsed across feedback delay). Score values ranged from 2.47 (for shaken oui) to 3.51 (for neutral oui) for female listeners, and from 2.51 (for questioning oui) to 3.31 (for shaken oui) for male speakers. As for reaction times, a significant three way interaction INTONATION x DELAY x LISTENER SEX was found [β = -0.54, SE = 0.25, t = -2.13, p = .03]. While for female speakers there was no difference in RT across different prosodies and delays, male speakers took more time to judge the shaken oui when it was presented after a long delay than after a short delay. This is illustrated in Figure 2 (bottom).

For the phase "Future Implications", we found a significant effect of DELAY for the neutral oui [β = 0.24, SE = 0.12, t = 2.02, p = .047], in that the mean score was slightly higher after a short (mean score = 3.25) than after a long delay (mean score = 3.05). No other effects were significant. There were no differences across the experimental factors in reaction times.

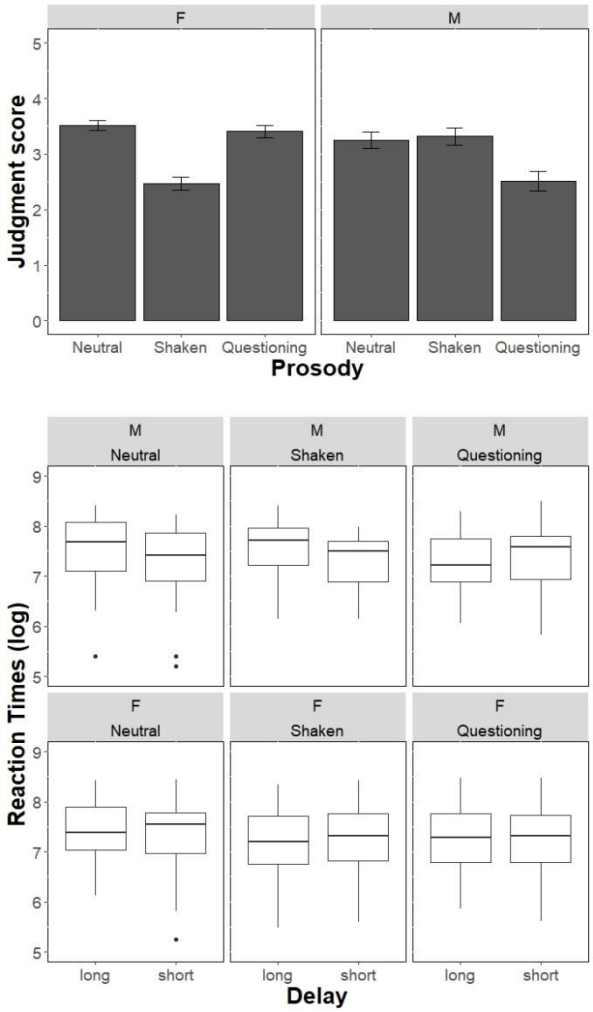


Figure 3. Judgment scores (top) and reaction times (bottom) within the context "Problem Definition"

4. Discussion

This study investigated the role of prosodic features in the perception of the linguistic feedback "oui" in a medical interactional context involving a patient's relative and a doctor. We focused on "oui" because it was very frequently found in the Acorformed corpus, especially in the two phases selected for the perception experiment, ie. Problem Definition and Future Implications. This suggested us it could be a good candidate for disambiguating between the different communicative functions that it could express according to its localization in the scenario (Problem Definition -PD- or Future Implications -FI-).

Means for judgment scores tended to be found in the middle of the rating scale range, suggesting that listeners interpreted the natural "oui" feedbacks as moderately appropriate to the previous context for both DP and IF. Despite this, prosody modulated the judgment score, with its effect being depending on listeners' sex. The shaken "oui" is judged as the less appropriate item by female listeners while the questioning oui is judged as the less appropriate item by male listeners. If the result for female listeners confirms our expectation, i.e. the most appropriate item in PD being a neutral or questioning item, the result concerning male listeners is quite novel. Indeed, the PD phase not only provides the opportunity to give feedback as continuer (explicit mark of listening and comprehension enabling the main speaker to continue his/her current discourse) but also questioning item that simply could signal a punctual trouble or misunderstanding. This PD phase is then more likely to be punctuated by a neutral or a questioning "oui" that a shaken one. This is confirmed for female listeners but not for male ones.

On the other hand, results of reaction times for male listeners suggest that, though male listeners do not find the shaken "oui" as an inappropriate item, they exhibit longer reaction time in their judgment when this same item is produced after a long delay. If we recall that PD rather tends to induce a short delay, this result suggests that male listeners are more sensitive to temporal delay than to the intonation of the feedback.

Finally, the FI phase does not provide significant results, except a very small effect on delay. This can be partly explained by the corpus study itself in which the distribution of "oui" items in FI were less sliced than in PD (larger acceptability for FI) maybe linked to a lack of data. Further investigations involving more data and also more participants will fill this gap and will allow us to deepen the gender difference that has emerged here.

In the context of the Acorformed project, the virtual patient's feedback is of importance given its principal role of listener. In order to improve the human-machine interaction, and in particular the perception of the virtual patient and the engagement of the user, the next step is to use the results of the study presented in this paper to model the prosodic features of the virtual patient's feedbacks. The implementation of the prosodic features in the virtual patient will enable us to explore, through a perception study in the context of human-machine interaction, the effect of the virtual appearance of the patient on the user's perception compared to the results obtain in this human-human perception study.

5. Acknowledgements

This work has been funded by the French National Research Agency project ACORFORMED (ANR-14-CE24-0034-02) and supported by grants ANR-16-CONV-0002 (ILCB), ANR-11- LABX-0036 (BLRI) and ANR-11-IDEX-0001-02 (A*MIDEX) Plus grant to Camilla Mastriani.

6. References

[1] J. Cassell (Ed.). (2000). *Embodied conversational agents*. MIT press. 2000.

[2] J. Gratch, A. Okhmatovskaia, F. Lamothe, S. Marsella, M. Morales, R.J. van der Werf, & L.P. Morency, "Virtual rapport", *International Workshop on Intelligent Virtual Agents*, 14-27, Springer Berlin Heidelberg. 2006.

[3] C. Porhet, M. Ochs, J. Saubesty, G. de Montcheuil, R. Bertrand, "Mining a Multimodal Corpus of Doctor's Training for Virtual Patient's Feedbacks", *International Conference on Multimodal Interaction (ICMI)*. 2017.

[4] R. Poppe, K.P. Truong, D. Reidsma, D. Heylen, "Backchannel strategies for artificial listeners", *International Conference on Intelligent Virtual Agents*, 146-158, Springer Berlin Heidelberg. 2010.

[5] Allwood, J., J. Nivre, and E. Ahlsen, "On the semantics and pragmatics of linguistic feedback", *Journal of Semantics*, 9 (1):1-30. 1992.

[6] **A.** Gravano, J. Hirschberg, S. Benus, "Affirmative cue words in task-oriented dialogue". *Computational Linguistics*. Vol. 38 (1), 1-39, 2012.

[7] Hockey, B. A, "Prosody and the role of 'okay' and 'uh-huh' in discourse", *Proceedings of the Eastern States Conference on Linguistics*, 128-136, Columbus, OH. 1993.

[8] J. Kowtko, *The Function of Intonation in Task-Oriented Dialogue*. Ph.D. thesis, University of Edinburgh. 1996.

[9] A. Pomerantz and J. Heritage, "Preference", In J. Sidnell & T.Stivers (eds), *The Handbook of Conversation Analysis*, Blackwell Publishing Ltd, 210-228. 2013.

[10] Stivers, T., Enfield, N.J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoyman, G., Rossano, F., de Ruiter, J.P. Yoon, K. & S.C. Levinson, "Universals and cultural variation in turn-taking in conversation", *Proceedings of the National Academy of Sciences, 106(26),* 10587-92. 2009.

[11] S.C. Levinson and F. Torreira, "Timing in turn-taking and its implications for processing models of language", In J. Holler, K.H. Kendrick, M. Casillas, M., S.C. Levinson eds. *Turn-Taking in Human Communicative Interaction*. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-825-2. 2016.

[12] R. Levitan, *Acoustic-prosodic entrainment in human-human and human-computer dialogue*. Ph.D. dissertation, Columbia University. 2014.

[13] R. Levitan, S. Benus, R. H. Galvez, A. Gravano, F. Savoretti, M. Trnka, A.Weise, J. Hischberg, "Implementing acoustic-prosodic entrainment in a conversational avatar", *Interspeech*, 1166-1170. 2016.

[14] L. Prevot, B. Bigi, and R. Bertrand, "A quantitative view of feedback lexical markers in conversational French", *14th Annual Meeting of the Special Interest Group on Discourse and Dialogue* , 1-4, 2013.

[15] C. André, A. Ghio, C. Cavé, B. Teston, "Perceval: a Computer-Driven System for Experimentation on Auditory and Visual Perception", *Proceedings of XVth ICPhS*, Barcelone, 1421-1424, 2003.

[16] J. D. Robinson, "An interactional structure of medical activities during acute visits and its implications for patients' participation", *Health Communication*, 15, 27-57, 2003.