Mining a Multimodal Corpus of Doctor's Training for Virtual Patient's Feedbacks

Chris Porhet, Magalie Ochs, Jorane Saubesty, Grégoire De Montcheuil, Roxane Bertrand

Mining a Multimodal Corpus of Doctor's Training for Virtual Patient's Feedbacks

Chris Porhet[2], Magalie Ochs[1], Jorane Saubesty[2], Grégoire de Montcheuil[2] and Roxanne Bertrand[2]
Aix-Marseille Université, CNRS, ENSAM, Université de Toulon, [1]LSIS UMR 7296, [2]LPL UMR 7309 ; France

ABSTRACT

Doctors should be trained not only to perform medical or surgical acts but also to develop competences in communication for their interaction with patients. For instance, the way doctors deliver bad news has a significant impact on the therapeutic process. In order to facilitate the doctors' training to break bad news, we aim at developing a virtual patient ables to interact in a multimodal way with doctors announcing an undesirable event. One of the key elements to create an engaging interaction is the feedbacks' behavior of the virtual character. In order to model the virtual patient's feedbacks in the context of breaking bad news, we have analyzed a corpus of real doctor's training. The verbal and nonverbal signals of both the doctors and the patients have been annotated. In order to identify the types of feedbacks and the elements that may elicit a feedback, we have explored the corpus based on sequences mining methods. Rules, that have been extracted from the corpus, enable us to determine when a virtual patient should express which feedbacks when a doctor announces a bad new.

Keywords

Embodied conversational agent, feedbacks, rules mining, multimodal corpora, health domain, virtual training.

ACM Reference Format

Chris Porhet, Magalie Ochs, Jorane Saubesty, Grégoire de Montcheuil, and Roxane Bertrand. 2017. Mining a Multimodal Corpus of Doctor's Training for Virtual Patient's Feedbacks. In Proceedings of 19th ACM International Conference on Multimodal Interaction (ICMI'17). ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3136755.3136816

1. INTRODUCTION

Doctors should be trained not only to perform medical or surgical acts but also to develop competences in communication for their interaction with patients. For instance, they often face the announcement of undesirable events to patients, as for example damage associated with care. A damage associated with care is the consequence of an unexpected event that can be due to complication connected to the pathology of the patient, unforeseeable medical situation, dysfunction or medical error. The damage may have physical, psychological, or even social and material repercussions.

The way doctors deliver bad news related to damage associated with care has a significant impact on the therapeutic process: disease evolution, adherence with treatment recommendations, litigation possibilities [1]. However, both experienced clinicians and medical trainees consider this task as difficult, daunting, and stressful. Nowadays, training health care professional to break bad news, recommended by the French National Authority for Health (HAS[1]), is organized as workshops during which doctors disclose bad news to actors playing the role of patients. In order to facilitate, the doctor's training, we aim at developing a virtual patient able to interact in natural language and in a multimodal way with doctors simulating breaking bad news situation with such a virtual patient. One key challenge is to create a believable virtual patient. In such a context of breaking bad news, the virtual patient has the main role of a listener in the conversation. Consequently, an important aspect is the production of feedbacks to create an engaging interaction. Indeed, in intelligent virtual agent domain, several researches have demonstrated the importance of embodied conversational agent's feedbacks for the interaction (e.g. perception of the agent, flow of the conversation, establishment of rapport) [16] and [29].

From a linguistic point of view, feedback responses (also called backchannel, Yngve [57] among others; for a more exhaustive list see [27]) are short verbal, vocal and gestural responses provided by recipient. As highlighted by Allwood in [2]: "The raison d'etre of linguistic feedback mechanisms is the need to elicit and give information about the basic communicative functions, i.e. continued contact, perception, understanding and emotional/attitudinal reaction, in a sufficiently unobtrusive way to allow communication to serve as an instrument for pursuing various human activities".

This project aims at improving the ability for virtual agents to produce a feedback in an appropriate way. In line with Ward & Tsukahara [56] or Gravano & Hischberg's studies [30] who attempted to characterize feedback/backchannel inviting-cues at a prosodic level, we propose to go beyond by attempting to identify sequences including some previously identified cues that may lead to the production of a feedback [6] at a verbal and gestural level. To our knowledge there is no study focusing on multimodal feedback inviting-cues in such an interactional setting.

In order to integrate feedbacks in a virtual patient in the specific context of breaking bad news, we have firstly adopted a corpus-based approach. Starting from a corpus of real training sessions in French medical institutions, we have performed multiple annotations at verbal and gestural levels, for both doctor and patient's speech (Section 3). Secondly, we have exploited sequences mining methods to extract from the corpus,

ICMI'17, November 13–17, 2017, Glasgow, UK

https://doi.org/10.1145/3136755.3136816

[1] The French National Authority for Health is an independent public scientific authority with an overall mission of contributing to the regulation of the healthcare system by improving health quality and efficiency.

rules on the patient's feedback elicitation (Section 4). Such a method has enabled us to identify the combinations of doctors' verbal and nonverbal signals that elicit different types of patient's feedback, verbal or nonverbal (Section 5).

2. RELATED WORKS

Several researches have been conducted to give the capabilities to Embodied Conversational Agents to express appropriate feedbacks during an interaction with a user. Gandalf [55] and REA [15] are examples of first virtual agents that integrate feedbacks as important features. Both of them express verbal and nonverbal feedbacks (e.g. "mhm" or head nodes) during their interaction with users. In the virtual agent Max [35], a behavior planning selects further non-verbal feedbacks (body gestures, head gestures, facial expressions) based on the conversational function.

Different approaches have been proposed to construct a model to predict the virtual agent's feedbacks during a conversation with a user. Based on research on human's feedbacks during interpersonal interactions, rule-based models have been defined. For instance, [8] and [46] predict virtual agent's feedbacks using speaker nods and speech cues. Another approach consists in exploiting machine learning algorithms to learn on a corpus when a virtual agent should display which feedbacks. For instance, based on a database of human-human interactions, [41] used Hidden Markov Model to predict listener backchannels using the speaker multimodal output (e.g., words, eye gaze and prosody). Even if machine learning approach has shown great potential, they present some limits. First, they require large corpora of annotated data to ensure the generalization of the model. Secondly, they remains "black box" since the results of such algorithms are difficult to understand and to represent explicitly. In our research, we have explored sequences mining algorithms. This choice is motivated by the fact that this kind of data mining algorithms enables us to extract explicit rules that are easily understood and interpretable in the light of existing research in linguistics. As far as we know, sequences mining algorithms have not been exploited to construct virtual agent's feedbacks models. However, these algorithms have been used to develop non-verbal behavior models of virtual agents. For instance, Chollet et al. [18] have used sequence-mining algorithm (the Generalized Sequence Pattern (GSP) algorithm [53]) in order to find simple sequences of non-verbal signals associated to social attitudes. Note that they have focused specifically on the nonverbal level. In [19] and [14], the authors have also applied a sequential pattern mining technique to automatically extract social signals (considering prosody, head movements and facial muscles) to characterize interpersonal attitudes. They particularly focus on the temporal characteristics of sequences (e.g. duration, delay). Whereas certain research works have explored sequences mining methods to extract information on the characteristics of user's behavior in corpora, one major limit of these works is to consider only one interlocutor, and not the signals expressed by one individual in response to the signals expressed by her interlocutor. Moreover, most of existing research works have focused on specific verbal or non-verbal signals. In the work presented in this article, we aim at analyzing the interaction of both verbal and non-verbal signals of the two interlocutors to model the feedbacks behavior.

3. MULTIMODAL CORPUS

In order to model the patient's feedbacks, we have studied an audio-visual corpus of interactions between doctors and actors playing the role of patients during real training sessions in French medical institutions. Indeed, for ethical reasons, it is not possible to videotape real breaking bad news situations. Instead, simulations are organized with actors playing the role of the patient (called "Standardized patients"). The use of "Standardized Patients" in medical training is a common and existing practice, to train doctor when the situations are sensitive. The actors are carefully trained and follow specific protocols define by experts. A corpus of such interactions has been collected in two different medical institutions. Simulated patients are actors trained to play the most frequently observed patients reactions. The actor follows a pre-determined scenario. The doctor (*i.e.* the trainee) receives details of a medical case before the simulated interaction starts (patient medical history, family background, surgery, diagnosis, etc.). On average, a simulated consultation lasts 15 minutes. The collected corpus is composed of 13 videos of patient-doctor interaction (each video involved a different doctor and, most of the time, a different actor-patient) with different scenarios. The total duration of the annotated corpus is almost 2 hours (119 minutes).

Different tools have been used in order to annotate the corpus. First, the corpus has been automatically segmented using SPPAS [10] and manually transcribed using PRAAT [11]. The doctors' and patients non-verbal behaviors have been manually annotated using ELAN [52].

Three annotators coded the corpus. Each of them annotated a third of the corpus. The annotators were graduate students in linguistics and were paid to annotate. In order to insure homogeneity among the annotators, a guide was given describing every annotation steps. Moreover, the annotations' sessions were supervised, allowing the annotators to ask question at any moment.

3.3 Verbal Cues Annotations

Audio files were extracted from the video recordings. The speech signal was segmented into Inter-Pausal Units (IPUs), defined as speech blocks surrounded by at least 200 ms silent pauses[2]. Due to its objective nature [36], the IPU can be automatically segmented. However, due to poor audio quality, they were manually corrected. We manually transcribed each participant's speech on two different tiers using the TOE convention (Transcription Orthographique Enrichie / Enriched Orthographical Transcription, [7]). Note that we do not consider the acoustic features (e.g. prosody) since the audio quality of the videos does not enable us to study this aspect. The part-of-speech (POS) tags were automatically identified using MarsaTag [49]. MarsaTag is a stochastic parser for written French which has been adapted to account for the specificities of spoken French. Among other outputs, it provides a morpho-syntactic category for each POS token.

3.4 Visual Cues Annotations

Different modalities of both the doctors and the patients have been annotated. The modalities as well as the corresponding

[2] For French language, lowering this 200 ms threshold would lead to many more errors due to the confusion of pause with the closure part of unvoiced consonants, or with constrictives produced with a very low energy.

values are described in Table 1[3].

Table 1: Non-verbal modalities

Modality	Values	Ref.
Head movements	nod, shake (negation), tilt, bottom, up, side	[4]
Posture change (movements of the bust)	forward, backwards, other change	[12]
Gaze direction	oneself, interlocutor, other direction, closed eyes	[7], [31]
Eyebrow expression	frown, raise	[28]
Hand gesture	movement[4]	[39]
Smile	smile, no smile[5]	[21]

We summarize the annotation for each interlocutor in Table 2. The table reveals that the most frequent non-verbal signals are the doctor's and patient's head movements while few smiles appear. The number of words shows that the doctors speak more than the patient, as expected given the context of the interaction.

In order to validate the annotation, 5% of the corpus was annotated by one more annotator. Few differences have been identified, differences mainly due to the quality of the video and the ability of the annotators to observe or not micro-movements. We calculated an inter-coder agreement using Cohen's Kappa. Agreement calculated was satisfying (k=0.63).

4. AUTOMATIC EXTRACTION OF MULTIMODAL SEQUENCES LEADING TO FEEDBACKS

In order to identify the doctor's verbal and nonverbal cues that lead to patients feedbacks, we have extracted, from the annotated corpus, sequences of multimodal signals ending by feedbacks. In a first step, we have defined a set of hypotheses in the light of research in linguistics in order to determine both the modalities to consider in the sequences and the segmentation of the sequences. In a second step, the set of sequences has been extracted using sequences mining algorithms (Section 4.1) and, based on these sequences, rules have been identified for feedback elicitation (Section 4.2).

4.1 Mining Multimodal Sequences Ending by Feedbacks

Based on several research on feedbacks both in linguistics and in the domain of virtual agent [20, 13, 35] and given the corpus, we have focused on the following feedback types: head movements, hand gestures, eyebrows movements, smiles, posture changes and gaze direction. Concerning the verbal feedbacks, we base our work on [47] that have constituted a list of French verbal expressions frequently used as feedback (e.g. "oui" (yes), "hmm", "euh", "d'accord" (ok), "hein", "non" (no), "ouais" (yea)).

[3] Note that we have not annotated the dynamic of the visual cues (e.g. amplitude). Because the manual annotation of the dynamic is really time-consuming, we aim at testing the impact of dynamics directly with the virtual patient that is easily controllable.

[4] As we are interested only in movements, we did not differentiated one movement from another. The hand annotation indicate the time interval from the moment the hands start moving until they return to the rest position.

[5] The video quality does not enable us to distinguish the different types of smile.

Table 2: Total number of annotations per interlocutor

Category	Number of annotations	
	doctors	patients
Head	3649	1970
Hands	635	463
Gaze	1823	716
Smile	20	20
Eyebrows	225	189
Posture	239	257
Words	44816	16727

To extract sequences from the multimodal corpora, we have defined how to segment the sequences, i.e. defined when a sequence leading to a feedback may start and end. The sequences have been segmented based on the IPUs (Section 3.3). The span of meaningful events (for the feedback production) is restrained to the current IPU. Given that we consider IPU as a form of turn-at-talk, we hypothesized that the end of an IPU could be a possible completion of a turn at which the interlocutor can provide a feedback response. However, IPU can be very long. So we chose to investigate the end of the IPU by limiting our observation to the last five tokens preceding the feedback. This choice is related to the size of units established for French spontaneous corpus in [48] (mean prosodic units). The IPU are then segmented in sequence when a non-verbal or verbal feedback occurs during or a short time after the IPU (~200ms) [54].

Following Bertrand et al. in [6] who have stated that feedback often occur more preferentially after content words such as nouns, verbs and adverb, we only retained these parts of speech. Moreover, we have added medical specific terms (identified based on a dictionary of French medical terms[6]). Note that we didn't consider the exact word used in the sequences mining but the category of the words (nouns, verbs, adverbs, medical terms).

Based on these hypotheses on the segmentation and modalities, we have developed a specific script to extract the corresponding sequences using the SPPAS' python API [10]. In total, we have extracted 8368 sequences from the corpus. Some examples of the obtained sequences are described in Table 3[7].

The extracted sequences show that most of the sequences seems to characterize head feedbacks whereas few smile feedbacks have been identified in the corpus, that can be easily explained by the context of the interaction (breaking bad news). Based on this corpus of sequences ending by feedbacks, the next step is to extract rules to determine, in fact, the doctor's verbal and nonverbal cues that elicit patient's feedbacks.

Table 3: Some examples of extracted sequences

doctor_adverb; doctor_verb; patient_head_shake;
doctor_gaze_interlocutor;doctor_hands_mvt;patient_hands_mvt
doctor_head_nod; doctor_noun;patient_hand_mvt;
doctor_gaze_interlocutor;patient_head_tilt
doctor_gaze_interlocutor;doctor_head_nod;patient_hands_mvt;
doctor_gaze_interlocutor;doctor_head_nod;patient_head_side;

[6] https://www.vocabulaire-medical.fr/

[7] All the extracted sequences as well as descriptive statistics are described at http://www.lpl-aix.fr/~acorformed/17-ICMI/annexe_1.html

4.2 Identify the Causal Relations between Multimodal Sequences and Feedback Responses

Our objective was to extract from the corpus of sequences described in the previous section a set of sequential rules that enables us to identify the causal relations between the verbal and nonverbal cues expressed by the doctors and the feedbacks of the patients. A sequential rule (also called episode rule, temporal rule or prediction rule) X⇒Y is a sequential relationship between two sets of items X and Y such that X and Y are disjoint, and all elements of X appeared before any elements of Y, and items of X and Y are unordered among themselves. Note that the fact that X and Y are unordered does not enable us to study the link between the order of the cues expressed by the doctor and the feedbacks. However, this choice is motivated by the fact that since X and Y are not time-ordered, we can make more accurate prediction as explained by Fournier in [22]. The quality of the sequential rules are generally characterised by the support and the confidence. The support of a rule X⇒Y is the number of sequences that contains X∪Y divided by the number of sequences in the database (that is how many times the sequence (X, Y) appears in the data). The confidence of a rule (in [0;1]) is the number of sequences that contains X∪Y, divided by the number of sequences that contains X (that is the proportion of cases that verify the rule: X and Y are both observed - among the cases containing the premises X). To extract sequential rules from the sequences corpus, we have used the algorithm ERMiner proposed by Fournier-Viger et al [26]. ERMiner is known as a very efficient algorithm, faster than another sequential rules algorithms such as CMDeo [23], CMRules [24], or RuleGrowth [25].

Table 4: Top 11 of the most confidante rules extracted

Rules	Conf.
doctor_hand_mvt ⇒ patient_gaze_interlocutor	0.36
doctor_head_nod ⇒ patient_head_nod	0.29
doctor_hand_mvt, doctor_verb, doctor_noun, doctor_gaze_interlocutor ⇒ patient_head_nod	0.25
doctor_hand_mvt, doctor_gaze_interlocutor ⇒ patient_gaze_interlocutor	0.25
doctor_verb, doctor_head]nod, doctor_noun ⇒ patient_head_nod	0.26
doctor_adverb ⇒ patient_head_nod	0.18
doctor_medical_vocabulairy ⇒ patient_head_nod	0.16
doctor_verb, doctor_adverb, doctor_gaze_interlocutor ⇒ patient_hand_movement	0.16
doctor_verb ⇒ patient_head_nod	0.17
doctor_gaze_interlocutor, doctor_head_shake ⇒ patient_hand_mvt	0.15
doctor_noun ⇒ patient_head_nod	0.12

The 10 rules with the highest confidence are described in Table 4[8]. For instance, the first rule may be interpreted as "a doctor hand movement may elicit a feedback of the patient corresponding to a gaze at the doctor". Note that the low confidence associated to rules can be explained by the fine-grained granularity of the rules. To obtain a better confidence, we are currently extracting rules considering, not the precise type of the feedback (such as head tilt) but the category (such as head movement) or, in a higher level, only the presence of a feedback whatever is the type or the category.

5. DISCUSSION

The results of the corpus mining show that nods, hand movements and gaze are the most frequent feedbacks compared to vocal feedbacks, smiles and posture that do not appear in the rules. Our results concerning head movements corroborate with preliminary results [50]. This confirms the literature [20], [5] and [43] in which head movements would be one of the main gestural cues to signal feedback. Nods and hand gestures could be then interpreted according to the different functions of feedback when implementing in the virtual agent: with nods as continuer or acknowledgement [51] the agent would provide explicit marks of listening and comprehension while not interrupting the main speaker during the first phases of explanation. With hand gestures as assessment [51] including news markers, the patient rather would express his/her stance (his/her trouble, anxiety) concerning the bad news announcement and its future implications.

Considering the most important feedback inviting-cues, we can see that hands movement (from doctor) lead to gaze (from patient), nods (from doctor) often lead to nods (from recipient) as well as gaze (from doctor) leads to gaze (from patient). This confirms Bertrand et al [6] who explained this type of pattern as the establishment of a communication mode in which a gesture answers another gesture. More specifically, such a matching (nod/nod; gaze/gaze) can reveal a form of alignment whereby participants make the success of the interaction easier [45].

In conclusion, in this paper, we have investigated the verbal and nonverbal cues that may elicit feedbacks' interlocutor in the specific context of a doctor announcing a bad new to a patient. Sequences mining methods have enabled us to identify rules on the multimodal triggers of different types of feedbacks, results supported by previous research in linguistics. The next steps are to pursue the study by analyzing more precisely the effects of the order of the signals and of their temporal characteristics on the types of feedback. To consider the impact of the flow of the dialog, we aim at coding "specific parts" in the dialog (e.g. reassuring part): one hypothesis is that the feedback behavior depends on theses dialog phases. Moreover, we are currently validating the rules in two ways: (1) in a classical machine learning perspective, we have divided the corpus in a learning set and a test set and through a k-cross validation, we are validating the rules learned on the learning set on the test set, (2) in a virtual agent perspective, we are integrating the rules in the virtual patient to measure the perception of users (through perceptive test to measure the perceptive believability). The rules and the validation remain specific to the studied context and should be evaluated in other contexts of interaction for generalization.

ACKNOWLEDGMENTS

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)

[8] Rules are described at www.lpl-aix.fr/~acorformed/17-ICMI/annexe_1.html

REFERENCES

[1] Andrade AD, Bagri A, Zaw K, Roos BA, Ruiz JG. Avatar-mediated training in the delivery of bad news in a virtual world. J Palliat Med 2010;13:1415-9.

[2] Allwood, J., Nivre J., Ahlsén E., (1992), On the semantics and pragmatics of linguistic feedback Journal of semantics 9 (1), 1-26,

[3] Allwood, J., Kopp, S., Grammer, K. et al. Lang Resources & Evaluation (2007) 41: 255.

[4] Allwood, J., & Cerrato, L. (2003). A study of gestural feedback expressions. In First nordic symposium on multimodal communication (pp. 7-22). Copenhagen.

[5] Battersby, S. A., and Healey, P. G., "Head and hand movements in the orchestration of dialogue.", in Annual Conference of the Cognitive Science Society, 2010

[6] Bertrand, R., Ferré, G., Blache, P. , Espesser, R., Rauzy, S.. Backchannels revisited from a multimodal perspective. Auditory-visual Speech Processing, Aug 2007, Hilvarenbeek, Netherlands. pp.1-5, (2007).

[7] Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B. and Rauzy, S. (2008). Le CID- Corpus of Interactional Data – Annotation et exploitation multimodale de parole conversationnelle. Traitement Automatique des Langes, 49(3), 1- 30.

[8] Bevacqua, E., Mancini, M., & Pelachaud, C. (2008, September). A listening agent exhibiting variable behaviour. In International Workshop on Intelligent Virtual Agents (pp. 262-269). Springer Berlin Heidelberg.

[9] Bevacqua, E., Hyniewska, S. J., & Pelachaud, C. (2010, May). Positive influence of smile backchannels in ECAs. In International Workshop on Interacting with ECAs as Virtual Characters (p. 13).

[10] Bigi, B. (2012). SPPAS: a tool for the phonetic segmentations of Speech. In The eighth international conference on Language Resources and Evaluation, 1748-1755.

[11] Boersma, P. P. G. (2002). Praat, a system for doing phonetics by computer. Glot international, 5.

[12] Bull, P.E. (1987) Posture and gesture. Oxford Pergamon Press.

[13] Brunner, L. J. (1979). Smiles can be back channels. Journal of Personality and Social Psychology, 37(5), 728.

[14] Cassell, J., & Thorisson, K. R. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. Applied Artificial Intelligence, 13(4-5), 519-538.

[15] Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H., Yan, H.: Embodiment in Conversational Interfaces: Rea. In: Proc. CHI, pp. 520–527 (1999)

[16] Chollet, M., Ochs, M., Clavel, C., & Pelachaud, C. (2013, September). A multimodal corpus approach to the design of virtual recruiters. In Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on (pp. 19-24). IEEE.

[17] Chollet, M., Ochs, M., & Pelachaud, C. (2014, May). Mining a multimodal corpus for non-verbal signals sequences conveying attitudes. In International Conference on Language Resources and Evaluation.

[18] Chollet, M., Ochs, M., & Pelachaud, C. (2014, August). From non-verbal signals sequence mining to bayesian networks for interpersonal attitudes expression. In International Conference on Intelligent Virtual Agents (pp. 120-133). Springer International Publishing.

[19] Dermouche, S., & Pelachaud, C. (2016, October). Sequence-based multimodal behavior modeling for social agents. In Proceedings of the 18th ACM International Conference on Multimodal Interaction (pp. 29-36). ACM.

[20] Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. Journal of personality and social psychology, 23(2), 283.

[21] Ekman, P., and Friesen, W. V. (1982). Felt, false, and miserable smiles. Journal of nonverbal behavior, 6(4), 238-252.

[22] Fournier-Viger, P. Gueniche, T., Tseng, V.S.: Using Partially-Ordered Sequential Rules to Generate More Accurate Sequence Prediction. Proc. 8th International Conference on Advanced Data Mining and Applications, pp. 431-442, Springer (2012)

[23] Fournier-Viger, P., Faghihi, U., Nkambou, R., Mephu Nguifo, E. (2012). CMRules: Mining Sequential Rules Common to Several Sequences. Knowledge-based Systems, Elsevier, 25(1): 63-76.

[24] Fournier-Viger, P., Faghihi, U., Nkambou, R., Mephu Nguifo, E. (2012). CMRules: Mining Sequential Rules Common to Several Sequences. Knowledge-based Systems, Elsevier, 25(1): 63-76.

[25] Fournier-Viger, P., Nkambou, R. & Tseng, V. S. (2011). RuleGrowth: Mining Sequential Rules Common to Several Sequences by Pattern-Growth. Proceedings of the 26th Symposium on Applied Computing (ACM SAC 2011). ACM Press, pp. 954-959.

[26] Fournier-Viger P., Gueniche T., Zida S., Tseng V.S. (2014) ERMiner: Sequential Rule Mining Using Equivalence Classes. In: Blockeel H., van Leeuwen M., Vinciotti V. (eds) Advances in Intelligent Data Analysis

[27] Gardner R., 2001. When listeners talk. John Benjamins Company.

[28] Goujon, A., Bertrand, R., and Tellier, M. (2015). Eyebrows in French talk-in-interaction. Gesture and Speech in Interaction, Sep 2015, Nantes, France.

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

[30] Gravano A., Hirschberg J., 2011. "Turn-taking cues in task-oriented dialogue". Computer Speech and Language 25(3), 601-634.

[31] Gullberg, M., and Kita, S. (2009). Attention to speech-accompanying gestures: Eye movements and information uptake. Journal of nonverbal behavior, 33(4), 251-277.

[32] Janssoone, T., Clavel, C., Bailly, K. and Richard, G., 2016, September. Using temporal association rules for the synthesis of embodied conversational agents with a specific stance. In International Conference on Intelligent Virtual Agents (pp. 175-189). Springer International Publishing.

[33] Jokinen, K.: Non-verbal Feedback in Interactions. In: Tao, J.H., Tan, T.N. (eds.) Affective Information Processing, pp. 227–240. Science+Business Media LLC. Springer, London (2008)

[34] Jokinen, K. (2009). Gaze and gesture activity in communication. Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments, 537-546.

[35] Kopp S., Allwood J., Grammer K., Ahlsen E., Stocksmeier T. (2008) Modeling Embodied Feedback with Virtual Humans. In: Wachsmuth I., Knoblich G. (eds) Modeling Communication with Robots and Virtual Humans. Lecture Notes in Computer Science, vol 4930. Springer, Berlin, Heidelberg

[36] Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A. and Den, Y. (1998). An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. Language and speech, 41(3-4), 295-321.

[37] Levitan, R., & Hirschberg, J. (2011, August). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In Interspeech (pp. 3081-3084).

[38] Martínez, H. P., & Yannakakis, G. N. (2011, November). Mining multimodal sequential patterns: a case study on affect detection. In Proceedings of the 13th international conference on multimodal interfaces (pp. 3-10). ACM.

[39] McNeill, D. (1992). Hand and mind: What gestures reveal about thought. University of Chicago press.

[40] Mota, S., and Picard, R. W. (2003). Automated posture analysis for detecting learner's interest level. In Computer Vision and Pattern Recognition Workshop, 2003. CVPR 03. Conference on (Vol. 5, pp. 49-49). IEEE.

[41] Morency, L. P., de Kok, I., & Gratch, J. (2010). A probabilistic multimodal approach for predicting listener backchannels. Autonomous Agents and Multi-Agent Systems, 20(1), 70-84.

[42] Niewiadomski R., Bevacqua E., Mancini M., Pelachaud C. Greta: an interactive expressive ECA system, Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009), Decker, Sichman, Sierra, and Castelfranchi (eds.), May, 10–15., 2009, Budapest, Hungary.

[43] Paggio, P., and Navarretta, C., "Feedback in head gestures and speech", in LREC 2010 Workshop Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality, 1-4,2010.
[44] Philips, E. (1999). The classification of smile patterns. J Can Dent Assoc, 65, 252-4.
[45] Pickering M.J. and Garrod S. 2006 Alignment as the Basis for Successful Communication, Research on Language and Computation, Springer, 203-228
[46] Poppe, R., Truong, K. P., Reidsma, D., & Heylen, D. (2010, September). Backchannel strategies for artificial listeners. In International Conference on Intelligent Virtual Agents (pp. 146-158). Springer Berlin Heidelberg.
[47] Prévot, L., Bigi, B., & Bertrand, R. (2013, August). A quantitative view of feedback lexical markers in conversational French. In 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue (pp. 1-4).
[48] Prevot, L., Gorish, J., & Mukherjee, S. (2015, October). Annotation and classification of french feedback communicative functions. In The 29th Pacific Asia Conference on Language, Information and Computation.
[49] Rauzy S., Montcheuil G., Blache P., 2014. "MarsaTag, a tagger for French written texts and speech transcriptions". Second Asia Pacific Corpus Linguistics Conference (2014 March 7-9: Hong Kong).
[50] Saubesty, J., & Tellier, M. (2016, September). Multimodal analysis of hand gesture back-channel feedback. In Gesture and Speech in Interaction 4 (pp. 205-210).
[51] Schegloff E.A., 1982. "Discourse as an interactional achievement: some uses of "uhhuh" and other things that come between sentences". In Tannen (ed.) Analyzing Discourse:Text and Talk. Washington, DC: Georgetown University press, 71-93
[52] Sloetjes, H., and Wittenburg, P. (2008). Annotation by Category: ELAN and ISO DCR. In LREC.
[53] Srikant, R., & Agrawal, R. (1996). Mining sequential patterns: Generalizations and performance improvements. Advances in Database Technology—EDBT'96, 1-17.
[54] Tellier, M., Guardiola, M., & Bigi, B. (2011). Types de gestes et utilisation de l'espace gestuel dans une description spatiale: méthodologie de l'annotation. Actes du 1er DEfi Geste Langue des Signes (DEGELS 2011), 45-56.
[55] Thórisson, K.R.: Communicative Humanoids: A Computational Model of Psychosocial Dialogue Skills. Ph.D. thesis, MIT (1996)
[56] Ward, N., & Tsukahara, W. (2000). Prosodic features which cue back-channel responses in English and Japanese. Journal of pragmatics, 32(8), 1177-1207.
[57] Yngve, V. H. (1970). On getting a word in edgewise. In Chicago Linguistics Society, 6th Meeting (pp. 567-578).