Annotating and Analyzing Emotions in a Corpus of First Encounters

Costanza Navarretta
Centre for Language Technology
University of Copenhagen
Njalsgade 140, build. 25 4th floor
2300 Copenhagen S, Denmark
Email: costanza@hum.ku.dk

***Abstract*—The paper is about the annotation and analysis of emotions and attitudes in a Danish corpus of first encounters. Emotions are classified using an open-ended list of values following the MUMIN annotation scheme and bipolar values in three emotional dimensions: Pleasure (P), Arousal (A) and Dominance (D), PAD) as proposed by Kipp and Martin [1]. The strategy of combining a coarse grained dimensional annotation with a fine grained list of emotions helped the coders to reach a common understanding of the semantics of the emotion labels. It also contributed to a significant improvement of the inter-coder agreement. A first analysis of the emotions annotated in six first encounters indicates that there is a tight relation between type of emotions and the social activity. Furthermore, some emotion types are related to specific communicative functions, such as feedback and turn management in these data. These findings can be useful for modeling plausible emotional devices.**

I. Introduction

This paper is about the annotation and analysis of emotions, attitudes and affective states in a Danish corpus of video recorded dyadic conversations between young people who do not know each other in advance [2]. First encounters have been analyzed in intercultural studies, inter alia [3], [4] because they can reveal how different cultures deal with varying degrees of familiarity as well as with social status and norms. Thus we investigate intra-cognitive communication [5].

Emotion studies involve numerous disciplines, such as psychology, sociology, linguistics and computer science. The point of view and the goal of these studies vary. Our research field is language technology and we study human-human multimodal communication, thus we are interested in individuating, analyzing and processing verbal and non-verbal behaviors in human communication. More specifically, in this paper we focus on the individuation, annotation and analysis of emotions, attitudes and affective states in video-recorded first encounters. Annotating emotions and affective states in a reliable way and analyzing their occurrences in different communicative situations and social activities do not only contribute to the implementation of plausible emotional devices, but also provide data for comparing intra- and inter-communication.

Many researchers have focused on the study of the six basic emotions: anger, happiness, sadness, disgust, fear and surprise starting with [6]. However, other emotions, attitudes and affective states are more relevant in everyday conversations, although they are often harder to identify than the six basic emotions.

In the following, *emotion* is used as a general term for emotion, attitude and affective states.

Interest in how emotions are expressed in communication is increasing because of the potential of including emotional factors in human-machine interfaces. Emotion-oriented computing comprises the recognition and generation of emotions. Recognition studies have both investigated emotions in single modalities, such as speech [7], [8], music [9], facial expressions [10], [11], body movements [12], [13] and hand gestures [1], and in more modalities [14]–[17]. Research on emotion generation has inter alia provided embodied conversational agents with basic emotions and different personality traits [18], [19].

Modeling the use of emotions in communication is a complex task also because it depends on numerous factors including the communicative situation, the relation between the participants, the social activity and the topics of the conversations. Annotators can also interpret emotions differently, and the situation is complicated by the fact that emotion labels are often ambiguous, see inter alia [20]. Thus, it is difficult to identify and classify emotions in a reliable way.

Various models of emotions have been proposed in different disciplines, and they fall into two main groups [20]: i) discrete categorial models using a finite list of independent emotions such as those proposed in [6], [21]; and ii) dimensional models of various complexity in which emotions are described by their location on the dimensions' axes. The first of these models was introduced by Wundt [22]. A semantic grid of emotion labels with respect to more dimensions has been proposed by Scherer [20], while a list of 48 common emotion labels has been individuated in various models, and is used in the Humaine net[1].

Since only some of the emotions listed in preceding studies seem to be relevant to the corpus of first encounters, and since preceding annotation work has shown that it is difficult for more annotators to get a common understanding of emotion labels, we wanted to test both a dimensional and a discrete

[1]Humaine net http://emotion-research.net.

978-1-4673-5188-1/12/$31.00 ©2012 IEEE

emotion model before deciding how to annotate emotions in the first encounter corpus.

In the first study, four coders annotated independently emotion labels on facial expressions in a first encounter conversation according to a coarse grained dimensional model proposed by Kipp and Martin [1]. In the second study the annotators used a fine grained and open-ended list of emotions as defined by the MUMIN annotation scheme [23]. The inter-coder agreement in both cases was under 0.30. Two of the coders preferred the categorial annotation model and two other preferred the dimensional annotation model. All four coders recognized that the two systems had both advantages and disadvantages, thus we decided to combine the two systems assuming that they could complement each other. The combined annotation system has contributed to a significant improvement of inter-coder agreement scores.

In the paper we describe the emotion annotations of the corpus and the evaluation of these annotations, and present a preliminary analysis of the encodings in the first half of the corpus. The results of the study indicate that there is a correlation between communicative functions and emotions, and between the communicative situation and the emotions recognized in the corpus, thus studying emotion expressions in various communicative settings and social activities is necessary to be able to design cognitively plausible software agents.

In section II, we shortly present the corpus and the existing annotations. In section III, we describe the annotation of emotions and the inter-coder agreement tests which we performed. In section IV,a first analysis of the annotated emotions is presented. Finally, we conclude and present future work in section V.

II. The corpus

The Danish corpus of first encounters consists of twelve conversations between two subjects who do not know each other in advance [2]. It was collected at the University of Copenhagen under the Nordic NOMCO project, in which comparable multimodal corpora had to be collected and annotated in Danish, Finnish and Swedish in order to investigate specific communicative functions in the three countries [24]. The participants in the Danish corpus are aged between 20 and 36 years, and they are university students or university educated. The conversations were video recorded by three cameras in a studio and the participants were standing in front of each other and talked freely while getting acquainted. The subjects (6 males and 6 females) were involved in two conversations each, one with a man and one with a woman. A snapshot from the corpus is in figure 1.

Fig. 1. Snapshot from the corpus

TABLE I
Features describing facial expressions

Attribute	Value
General face	Smile, Laugh, Scowl, FaceOther
Eyebrows	Frown, Raise, BrowsOther
FeedbackBasic	CPU, SelfFeedback, FeedbackOther
FeedbackDirection	FeedbackGive, FeedbackElicit
FeedbackAgree	FeedbackAgree, FeedbackDisagree
TurnManagement	TurnTake,TurnElicit, TurnGive, TurnYield, TurnHold

The corpus contains orthographically transcriptions with word time stamps and annotations of communicative head movements, facial expressions and body expressions in the Anvil tool [25]. They are annotated following the MUMIN annotation scheme [23] which describes the shape, the semiotic types [26] and the communicative functions of these behaviors with pre-defined features. Co-speech bodily behaviors were also linked to words (own words or the interlocutor's ones) if the coders judged that they were semantically related. The bodily behaviors which are relevant for the present study are facial expressions and eyebrows, and the communicative functions which we include in our analysis are related to own communicative management (self-feedback) or interactive communication management (feedback and turn management) [27].

Table I contains an overview of the features relevant to this study. Facial expressions are described with a general face attribute and with a description of the movement of the eyebrows. The communicative function of feedback is described with three attributes, FeedbackBasic, FeedbackDirection and FeedbackAgreement. FeedbackBasic is assigned if feedback expresses a) Contact, Perception and Understanding (CPU), b) feedback by the speaker to her own speech contribution (SelfFeedback), c) or something else (FeedbackOther), that is simple Contact or Contact and Perception. FeedbackDirection indicates whether feedback is given or elicited, and FeedbackAgreement describes whether the gesturer is agreeing or disagreeing with the interlocutor.

The TurnManagement attribute is assigned when the speaker takes a turn that wasn't offered, possibly by interrupting (TurnTake); when the speaker wishes to keep the turn (TurnHold); when the speaker accepts a turn that is being offered (TurnAccept); when the speaker releases the turn under pressure TurnYield); and when the speaker offers the turn to the interlocutor (TurnElicit).

A more detailed description of the corpus and of the project in which it was annotated is in [24], [28].

III. The annotation of emotions

Emotions are not located in a single part of the body, but more modalities express them. We started coding emotions on facial expressions because it is assumed that facial expressions, together with verbal behavior, are very strong indicators of emotions. The coders were instructed to take into account the context when annotating the emotions expressed by the participants' face. This implies both the semantic content of the conversations and the other co-occurring modalities.

In the MUMIN annotation scheme, emotion encodings can be added to each modality. In the original MUMIN scheme [23] a list of approximately 20 emotions has been proposed, comprising the six basic emotions. The list is open-ended (there is a value EmotionOther), and more emotions can be added to the list if necessary. The fact that it is difficult to identify emotions and to reach agreement on emotion labels is recognized in the literature, inter alia [20]. Thus, we decided to test whether a coarse-grained dimensional system for annotating emotions [1] resulted in more reliable annotation than simply using emotion labels.

Kipp and Martin [1] annotate emotions using Russel and Mehrabian's three dimensional model [29]. In this model emotions are described according to their position in a three dimensional space. Thus the emotional level on the three dimensions Pleasure (P), Arousal (A) and Dominance (D) must be indicated. This three dimensional model is also called the PAD emotion space. In the model, the Pleasure dimension indicates positive versus negative affective state, the Arousal dimension indicates high versus low level of physical activation and/or mental alertness, and the Dominance dimension expresses the feeling of having control and influence over others and situations versus the feeling of being controlled and influenced by other or by the situation. In [29], Russel and Mehrabian asked 300 undergraduate students to place 151 emotions on the three axes. Kipp and Martin, instead, assume that each dimension is simply bipolar, thus only + and - values are allowed in their annotation. This results in eight PAD combinations.

In our pilot study, four coders had to annotate independently emotion labels on the facial expressions of a speaker, and then they had to add PAD labels on the facial expressions of the same speaker, but in another conversation. All four coders were experienced in annotating multimodal behaviors, but none of them had previously worked with emotions. Before starting annotating, they all received a short introduction to the PAD system, and they discussed the meaning of the existing emotion labels. They were also instructed to only code emotions on facial expressions if they meant that there was an emotion, and they could distinguish between spontaneous and acted emotions. The latter type is defined as a *Display*, which is a subtype of the semiotic class IndexicalNonDeictic [26].

Inter-coder agreement was calculated between pairs of coders in terms of Cohen's kappa [30]. Agreement was highest on the Pleasure and Dominance attributes, and lowest on the Arousal dimension. Two annotators had the best agreement scores (0.25 in average) while the annotators who disagreed mostly had scores between 0.15 and 0.20 on most categories. The disagreement on the 20 emotion labels was generally worse than on the PAD values, but two coders had higher agreement scores on most emotion attributes than on the PAD values. All coders meant that the three PAD dimensions were a useful way to individuate emotions, but that assigning bipolar values to the three dimensions, as suggested by Kipp and Martin [1] was problematic because the system did not allow distinguishing between similar emotions. In fact many emotions had the same PAD values, although they differ in degree of intensity along one or two dimensions. Furthermore, the coders used emotion labels when discussing disagreement cases.

In order to get the best from both dimensional and discrete models, we decided to combine each emotion label with its PAD combination. Then, two of the annotators made an annotation manual where emotions with the same PAD values were distinguished on the basis of their intensity on a 4 valued scale on each dimension [31]. This is a simplification of Russel and Mehrabian's method [29] which provides 18 different positions on each dimension.It must also be noted that some of the emotions which our coders added to the list do not occur in any of the discrete models we have found in the literature.

Two new inter-coder agreement experiments were performed with the two coders who disagreed mostly in the first experiment. The first test was performed after the emotion list had been merged with the PAD values, and the second test was completed after the two coders had assigned intensity scores to similar emotions in order to be able to distinguish them in a clear way. In both experiments they coded independently the emotions in a new conversation.

In the first experiment, the inter-coder agreement score was 0.39 [30] when distinguishing between 26 emotions. This is an improvement of 0.19 with respect to the earlier experiment where only emotion labels were used. In the second experiment the agreement score was 0.61. Scores between 0.60 and 0.70 are considered reliable on these types of annotation, see inter alia [32]. The log files showed that in 30% of the disagreement cases the annotators wrote that the facial expression indicated more than one emotion, e.g. the gesturer was both amused and friendly. Disagreement often involved emotions with the same PAD value, although this was not always the case. The last type of disagreement concerned deciding whether a facial expression expressed an emotion or not.

The improvement of inter-coder agreement scores from the very first test (annotation with emotion labels only) to the last test can be partly due to the fact that the coders got more experienced in annotating emotions and in the annotation scheme. Furthermore, they also came nearer to each other because they had discussed disagreement cases. However, the improvement between the two last experiments is big, and it was performed immediately after the preceding one. Both coders reported that they felt much more confident about the semantic differences of similar emotion labels after having placed them in the PAD dimension space.

After these experiments, the emotion coding procedure has been the following: one annotator codes the emotions expressed by facial expressions of one participant in a conversation, and a second annotator revises the annotations. An

TABLE II
LIST OF EMOTIONS AND THEIR PAD VALUES

P	A	D	Emotions
+	+	+	Amused, Excited, Happy, Ironic, Interested,Joking, Proud, Satisfied, Self-Confident,Supportive
-	+	-	Awkward, Embarrassed, Puzzled, Uncertain
+	-	+	Certain, Friendly
-	-	-	Disappointed, Hesitant, Uncomfortable, Unconfident, Uninterested
+	-	-	Docile, Thoughtful
+	+	-	Engaged, Surprised
-	+	+	Irritated

TABLE III
MOST FREQUENTLY OCCURRING EMOTIONS

Emotion	P	A	D	Number
Amused	+	+	+	119
Friendly	+	+	+	78
Interested	+	+	+	58
Hesitant	-	-	-	24
Certain	+	-	+	22
Supportive	+	+	+	21
Uncertain	-	+	-	17
Unconfortable	-	-	-	12
Satisfied	+	+	+	11
SelfConfident	+	+	+	11
Puzzled	-	+	-	10
Engaged	+	+	-	9
Surprised	+	+	-	8

agreed upon version is made. If disagreement persists, the final decision is taken by a third annotator. The roles between the two main annotators are swapped when annotating the emotions of the second participant.

New emotions are added to the list if all three annotators agree that the addition is necessary. The PAD encoding and the corresponding list of emotions which were recognized in the first six first encounters are in Table II.

Only one of the basic emotions (Happiness) has been recognized in the first encounters. The emotions which occurred in the corpus are clearly related to the type of conversation and social activity, as well as to the relation between the participants and their degree of familiarity.

Most of the emotions recognized in the first encounters concern the gesturer's reaction to the content of the conversations (both own and the other's contribution) and to the communicative situation. This is not surprising given the social activity the participants are involved in. They met a new person and must get acquainted in the course of the conversation. No emotion with negative valence for Pleasure and Arousal and positive valence for Dominance were found in the data.

IV. THE ANALYSIS OF THE EMOTIONS

448 out of the 642 facial expressions in the first six encounters have been judged to express an emotion (70% of the occurrences). The most frequently assigned emotion labels are in Table III.

TABLE IV
FREQUENCY OF PAD VALUES

PAD-value	Number
+++	240
+-+	100
—	44
-+-	32
++-	16
-++	8
+–	8

TABLE V
EMOTIONS AND INTERACTIVE FUNCTIONS

Function	Emotion	N
FeedbackGive	Friendly	48
	Amused	45
	Interested	38
	Supportive	19
SelfFeedback	Amused	65
	Hesitant	12
	Uncertain	9
	SelfConfident	7

Table IV shows the number of emotions with the same PAD value in the annotated data.

The table indicates that the most frequently occurring emotions in the corpus of first encounters have positive valence on all three axes. The second most frequently occurring emotion group has positive valence on the Pleasure and Dominance axes and negative valence on the Arousal axis, while the third group includes emotions with negative valence on all axes. Emotions with negative valence on the Pleasure and Dominance axes and positive valence on the Arousal axis are also frequent.

The fact that emotions with positive PAD valence are so frequent in this corpus is related to the social activity: the subjects meet for the first time, thus they are kind, they smile a lot, and show interest in what the interlocutor is saying.

We expected many emotional expressions to occur in relation with feedback-giving (giving feedback to the addressee's contribution) as well as showing attitudes to own contribution. The analysis of data confirms this as it can be seen in Table V. Here we show the emotions that occur most frequently in these two communicative contexts.

We also expected that some emotion labels would be occurring very frequently with Turn Management functions. More specifically, we expected that emotions such as Interested, Supportive and Friendly would be connected to the function of TurnElicit, and that the emotion Hesitant would be often connected to TurnHold. Finally we predicted that emotions classified with labels SelfConfident and Irritated often occurred with TurnYield. The analysis of the data seems to support the first two assumptions, but not the last one. TurnYield related facial expressions are not expressing any emotion in these data. However, the facial expressions related to turn management are too few to conclude on this.

A. Discussion

It is not surprising that the most frequently occurring emotions in the first encounters have a positive valence along all three or at least two dimensions. We also expected that they occurred in relation to the communicative functions of giving feedback to the addressee or of selffeedback. However, we do not know to which extent the degree of familiarity between the participants influences the expression of emotions. Familiarity has been recognized to influence the fluency of speech [33] and the frequency of feedback-related gestures [34], but it is not proved that people who met for the first time are friendlier and more smiling that people that know each other well. Furthermore, we expected that emotions such Awkward and Embarassed would occur more often in the data than they actually did, because the communicative situation is not completely natural (the conversations are recorded in a studio) and because we expected that some participants would be shy when interacting with an unknown person. But, this was not the case and seems to confirm the responses that the participants gave to a questionnaire about how they felt about the conversation after each recording [28].

In the analysis, we have exclusively looked at the frequency of the emotions visible in facial expressions in six conversations, but the frequency of some of the emotions varies from conversation to conversation. Thus, a more fine-grained analysis of these differences and of individual variation should be made.

In a preceding study [35] we found that there are gender related differences in the way the participants interact in this corpus. More precisely, male participants talk less (utter less words and keep the turn for shorter time) when interacting with a subject of the same gender, while they talk more while interacting with a woman. On the contrary, women talk more when interacting with another woman than with a man. Whether participants also show different emotions when talking with a male or a female will be investigated in future.

Cognitive studies suggest that there is a mirroring effect both when gesturing and when showing emotions [36], [37]. A first analysis of the data, also indicates that the same emotions are expressed by both subjects in a conversation, but to which extent this depends on a mirroring effect or on the content of the conversation should be analyzed further. Finally, even if the various facial expressions are linked to speech in the annotations, and the coders annotated emotions considering the context in which the emotion was expressed, we have not yet analysed the relation between the semantics of speech and the co-occurring emotions, neither we have annotated emotions in speech and in the other co-occurring bodily behaviors. Thus, the presented analysis is only a very preliminary study of the emotions in the first encounters corpus.

V. Conclusion

We have described the annotation work of emotions in a Danish corpus of first encounters according to an annotation scheme that combines a traditional discrete model of emotion labels with a simplification of Russel and Mehrabian's three dimensional PAD model [29]. The idea of using bipolar values for describing the emotions' position in the model is taken from Kipp and Martin [1].

Combining the two models, and discussing emotion labels in terms of a simplified PAD emotion space helped the annotators to get a common understanding of the emotion labels and contributed to a significant improvement of the inter-coder agreement scores. This is particularly important when annotating a corpus where most occurring emotions are similar and have the same PAD values.

A preliminary analysis of the annotations in six first encounters shows that the most frequently occurring emotions in this corpus have either positive valence along all the three dimensions Pleasure, Arousal and Dominance, or along the Pleasure and Dominance dimensions. This is clearly related to the social activity in which the participants were involved. Even if the conversation were video recorded in a studio, emotion labels such as Embarassed and Awkward were not frequently assigned which indicates that the participants felt the situation quite natural, and that they enjoyed the conversation.

The annotation and analysis of emotions and affective states in various types of conversation can contribute to the modelling of sophisticated emotional devices. Moreover, these data can be used to compare the expression of emotions in first encounters between humans and in first time interactions between humans and emotional devices.

In future, we will analyze both individual variation in the expression of emotions and the relation between the semantic content of the conversations on the one hand and emotions as expressed in both verbal and non verbal behaviors on the other.

Acknowledgment

The work described in this paper is funded by the NORDCORP program under the Nordic Research Councils for the Humanities and the Social Sciences (NOS-HS) and by the Danish Research Council for the Humanities (FKK). I would like to thank Anette Luft Studsgrd, Sara Andersen, Bjrn Wessel-Tolvig and Magdalena Lis who have coded the emotions with me. Finally, special thanks go to the NOMCO project's partners, Elisabeth Ahlsn, Jens Allwood, Kristiina Jokinen and, last but not least, Patrizia Paggio for the many interesting discussions and suggestions.

References

[1] M. Kipp and J.-C. Martin, "Gesture and emotion: Can basic gestural form features discriminate emotions?" in *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII-09)*, 2009. IEEE Press, 2009.

[2] P. Paggio and C. Navarretta, "Head movements, facial expressions and feedback in danish first encounters interactions: A culture-specific analysis," in *Universal Access in Human-Computer Interaction- Users Diversity. 6th International Conference. UAHCI 2011, Held as Part of HCI International 2011*, ser. LNCS, C. Stephanidis, Ed., no. 6766. Orlando Florida: Springer Verlag, 2011, pp. 583–690.

[3] M. Argyle, *Bodily Communication*. Methen & Co. Ltd, 1975.

[4] M. Rehm, E. Andre, N. B. amd Birgit Endrass, M. Wissner, Y. Nakano, A. A. Lipi, T. Nishida, and H.-H. Huang, "Creating standardized video recordings of multimodal interactions across cultures," in *Multimodal Corpora. From Models of Natural Interaction to Systems and Applications*, ser. LNAI, M. Kipp, J.-C. Martin, P. Paggio, and D. Heylen, Eds., no. 5509. Springer, 2009, pp. 138–159.

[5] P. Baranyi and A. Csapó, "Definition and synergies of cognitive infocommunications," *Acta Polytechnica Hungarica*, vol. 9, no. 1, pp. 67–83, 2012.

[6] P. Ekman and W. Friesen, *Unmasking the face. A guide to recognizing emotions from facial clues.* Englewood Cliffs, New Jersey: Prentice-Hall, 1975.

[7] R. Cowie and E. Douglas-Cowie, "Automatic statistical analysis of the signal and prosodic signs of emotion in speech," in *Proceedings of International Conference on Spoken Language Processing*, 1996, pp. 1989–1992.

[8] K. Scherer, "Adding the affective dimension: A new look in speech analysis and synthesis," in *Proceedings of International Conference on Spoken Language Processing*, 1996, pp. 1808–1811.

[9] T. Fritz, S. Jentschke, N. Gosselin, D. Sammler, I. Peretz, and R. Turn, "Universal recognition of three basic emotions in music," *Current Biology*, vol. 19, pp. 573–576, April 2009.

[10] M. Pantic and L. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424–1445, 2000.

[11] S. Ioannou, A. Raouzaiou, V. Tzouvaras, T. Mailis, K. Karpouzis, and S. Kollias, "Emotion recognition through facial expression analysis based on a neurofuzzy network," *Neural Networks*, vol. 18, no. 4, pp. 423–435, 2005.

[12] A. Camurri, I. Lagerlöf, and G. Volpe, "Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques," *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 213–225, 2003.

[13] G. Castellano, S. D. Villalba, and A. Camurri, "Recognising human emotions from body movement and gesture dynamics," in *ACII 2007*, ser. LNCS, A. Paiva, R. Prada, and R. Picard, Eds., no. 4738. Springer-Verlag, 2007, pp. 71–82.

[14] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. Lee, A. Kazemzaeh, S. Lee, U. Neumann, and S. Narayanan, "Analysis of emotion recognition using facial expressions, speech and multimodal information," in *Proceedings of ACM 6th international Conference on Multimodal Interfaces - ICMI 2004*, State College, PA, October 2004, pp. 205–211.

[15] G. Castellano, L. Kessous, and G. Caridakis, "Emotion recognition through multiple modalities: Face, body gesture, speech," in *Affect and Emotion in HCI*, ser. LNCS, C. Peter and R. Beale, Eds., no. 4868. Springer-Verlag, 2008, pp. 92–103.

[16] N. Bourbakis, A. Esposito, and D. Kavraki, "Extracting and associating meta-features for understanding people's emotional behaviour: Face and speech," *Cognitive Computation*, vol. 3, pp. 436–448, 2011.

[17] A. Esposito and M. T. Riviello, "The cross-modal and cross-cultural processing of affective information," in *Proceedings of the 2011 conference on Neural Nets WIRN10: Proceedings of the 20th Italian Workshop on Neural Nets*. Amsterdam, The Netherlands, The Netherlands: IOS Press, 2011, pp. 301–310.

[18] M. Ochs, D. Sadek, and C. Pelachaud, "A formal model of emotions for an empathic rational dialog agent," *International Journal of Autonomous Agents and Multi-Agent Systems*, 2010.

[19] R. Niewiadomski, S. Hyniewska, and C. Pelachaud, "Constraint-based model for synthesis of multimodal sequential expressions of emotions," *IEEE Transactions of Affective Computing*, vol. 2, no. 3, pp. 134–146, July 2011.

[20] K. R. Scherer, "What are emotions? and how can they be measured?" *Social Science Information*, vol. 44, no. 4, pp. 695–729, December 2005.

[21] A. Ortony, G. L. Clore, and A. Collins, *The cognitive structure of emotions*. Cambridge University Press, MA, 1988.

[22] W. Wundt, *Grundzge der physiologischen Psychologie*. Leipzig: Engelmann, 1905.

[23] J. Allwood, L. Cerrato, K. Jokinen, C. Navarretta, and P. Paggio, "The mumin coding scheme for the annotation of feedback, turn management and sequencing," *Multimodal Corpora for Modelling Human Multimodal Behaviour. Special Issue of the International Journal of Language Resources and Evaluation*, vol. 41, no. 3–4, pp. 273–287, 2007.

[24] C. Navarretta, E. Ahlsn, J. Allwood, K. Jokinen, and P. Paggio, "Feedback in nordic first-encounters: a comparative study," in *Proceedings of LREC 2012*, Istanbul Turkey, May 2012, pp. 2494–2499.

[25] M. Kipp, "Gesture generation by imitation - from human behavior to computer character animation," Ph.D. dissertation, Saarland University, Saarbruecken, Germany, Boca Raton, Florida, dissertation.com, 2004.

[26] C. S. Peirce, *Collected Papers of Charles Sanders Peirce, 1931-1958, 8 vols*. Cambridge, MA: Harvard University Press, 1931.

[27] J. Allwood, J. Nivre, and E. Ahls'en, "On the semantics and pragmatics of linguistic feedback," *Journal of Semantics*, vol. 9, pp. 1–26, 1992.

[28] P. Paggio and C. Navarretta, "Classifying the feedback function of head movements and face expressions," in *Proceedings of LREC 2012 Workshop Multimodal Corpora - How should multimodal corpora deal with the situation?,*, Istanbul Turkey, May 2012, pp. 2494–2499.

[29] J. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *Journal of Research in Personality*, vol. 11, pp. 273–294, 1977.

[30] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[31] A. L. Studsgaard and S. Andersen, "Annotating attitudes in the danish nomco corpus of first encounters," in *In extended abstracts of The Fourth Nordic Symposium on Multimodal Communication*, Gothenburg, Sweden, November 2012.

[32] T. Rietveld and R. v. Hout, *Statistical Techniques for the Study of Language and Language Behavior*. Berlin: Mouton De Gruyter, 1993.

[33] N. Campbell, "Individual traits of speaking style and speech rhythm in a spoken discourse," in *Proceedings of the COST 2102 Workshop*, 2007, pp. 107–120.

[34] C. Navarretta and P. Paggio, "Verbal and non-verbal feedback in different types of interactions," in *Proceedings of LREC 2012*, Istanbul Turkey, May 2012, pp. 2338–2342.

[35] Navarretta and Paggio, "Turn management in first encounters," in *In Abstracts of Workshop on The embodied foundation of human communicative skills*, Copenhagen, November 2012.

[36] G. Rizzolatti, "The mirror neuron system and its function in humans," *Anat.Embryol.*, vol. 210, pp. 419–421, 2005.

[37] G. Rizzolatti and M. Fabbri-Destro, "The mirror system and its role in social cognition," *Curr. Opin. Neurobiol.*, vol. 18, pp. 179–184, 2008.