Elisabeth André
Laila Dybkjær
Wolfgang Minker
Paul Heisterkamp (Eds.)

LNAI 3068

Affective Dialogue Systems

Tutorial and Research Workshop, ADS 2004
Kloster Irsee, Germany, June 2004
Proceedings

Springer

Elisabeth André Laila Dybkjær
Wolfgang Minker Paul Heisterkamp (Eds.)

Affective Dialogue Systems

Tutorial and Research Workshop, ADS 2004
Kloster Irsee, Germany, June 14-16, 2004
Proceedings

Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Elisabeth André
University of Augsburg, Laboratory for Multimedia Concepts and Applications
Eichleitnerstr. 30, 86135 Augsburg, Germany
E-mail: andre@informatik.uni-augsburg.de

Laila Dybkjær
University of Southern Denmark, Natural Interactive Systems Laboratory
Campusvej 55, 5230 Odense M, Denmark
E-mail: laila@nis.sdu.dk

Wolfgang Minker
University of Ulm, Department of Information Technology
Albert-Einstein-Allee 43, 89081 Ulm, Germany
E-mail: wolfgang.minker@e-technik.uni-ulm.de

Paul Heisterkamp
DaimlerChrysler AG, Dialog Systems, 89081 Ulm, Germany
E-mail: paul.heisterkamp@daimlerchrysler.com

Library of Congress Control Number: 2004106903

CR Subject Classification (1998): I.2, H.5.2-3, H.5.1, H.4, J.4, I.3.7

ISSN 0302-9743
ISBN 3-540-22143-3 Springer-Verlag Berlin Heidelberg New York

Springer-Verlag is a part of Springer Science+Business Media

springeronline.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP-Berlin, Protago-TeX-Production GmbH
Printed on acid-free paper SPIN: 11012603 06/3142 5 4 3 2 1 0

Lecture Notes in Artificial Intelligence 3068

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Springer
Berlin
Heidelberg
New York
Hong Kong
London
Milan
Paris
Tokyo

Lecture Notes in Artificial Intelligence (LNAI)

Vol. 3070: L. Rutkowski, J. Siekmann, R. Tadeusiewicz, L.A. Zadeh (Eds.), Artificial Intelligence and Soft Computing - ICAISC 2004. XXV, 1208 pages. 2004.

Vol. 3068: E. André, L. Dybkjær, W. Minker, P. Heisterkamp (Eds.), Affective Dialogue Systems. XII, 324 pages. 2004.

Vol. 3066: S. Tsumoto, R. S lowiński, J. Komorowski, J.W. Grzymala-Busse (Eds.), Rough Sets and Current Trends in Computing. XX, 853 pages. 2004.

Vol. 3065: A. Lomuscio, D. Nute (Eds.), Deontic Logic in Computer Science. X, 275 pages. 2004.

Vol. 3060: A.Y. Tawfik, S.D. Goodwin (Eds.), Advances in Artificial Intelligence. XIII, 582 pages. 2004.

Vol. 3056: H. Dai, R. Srikant, C. Zhang (Eds.), Advances in Knowledge Discovery and Data Mining. XIX, 713 pages. 2004.

Vol. 3035: M.A. Wimmer (Ed.), Knowledge Management in Electronic Government. XII, 326 pages. 2004.

Vol. 3034: J. Favela, E. Menasalvas, E. Chávez (Eds.), Advances in Web Intelligence. XIII, 227 pages. 2004.

Vol. 3030: P. Giorgini, B. Henderson-Sellers, M. Winikoff (Eds.), Agent-Oriented Information Systems. XIV, 207 pages. 2004.

Vol. 3029: B. Orchard, C. Yang, M. Ali (Eds.), Innovations in Applied Artificial Intelligence. XXI, 1272 pages. 2004.

Vol. 3025: G.A. Vouros, T. Panayiotopoulos (Eds.), Methods and Applications of Artificial Intelligence. XV, 546 pages. 2004.

Vol. 3012: K. Kurumatani, S.-H. Chen, A. Ohuchi (Eds.), Multi-Agnets for Mass User Support. X, 217 pages. 2004.

Vol. 3010: K.R. Apt, F. Fages, F. Rossi, P. Szeredi, J. Váncza (Eds.), Recent Advances in Constraints. VIII, 285 pages. 2004.

Vol. 2990: J. Leite, A. Omicini, L. Sterling, P. Torroni (Eds.), Declarative Agent Languages and Technologies. XII, 281 pages. 2004.

Vol. 2980: A. Blackwell, K. Marriott, A. Shimojima (Eds.), Diagrammatic Representation and Inference. XV, 448 pages. 2004.

Vol. 2977: G. Di Marzo Serugendo, A. Karageorgos, O.F. Rana, F. Zambonelli (Eds.), Engineering Self-Organising Systems. X, 299 pages. 2004.

Vol. 2972: R. Monroy, G. Arroyo-Figueroa, L.E. Sucar, H. Sossa (Eds.), MICAI 2004: Advances in Artificial Intelligence. XVII, 923 pages. 2004.

Vol. 2961: P. Eklund (Ed.), Concept Lattices. IX, 411 pages. 2004.

Vol. 2953: K. Konrad, Model Generation for Natural Language Interpretation and Analysis. XIII, 166 pages. 2004.

Vol. 2934: G. Lindemann, D. Moldt, M. Paolucci (Eds.), Regulated Agent-Based Social Systems. X, 301 pages. 2004.

Vol. 2930: F. Winkler (Ed.), Automated Deduction in Geometry. VII, 231 pages. 2004.

Vol. 2926: L. van Elst, V. Dignum, A. Abecker (Eds.), Agent-Mediated Knowledge Management. XI, 428 pages. 2004.

Vol. 2923: V. Lifschitz, I. Niemelä (Eds.), Logic Programming and Nonmonotonic Reasoning. IX, 365 pages. 2004.

Vol. 2915: A. Camurri, G. Volpe (Eds.), Gesture-Based Communication in Human-Computer Interaction. XIII, 558 pages. 2004.

Vol. 2913: T.M. Pinkston, V.K. Prasanna (Eds.), High Performance Computing - HiPC 2003. XX, 512 pages. 2003.

Vol. 2903: T.D. Gedeon, L.C.C. Fung (Eds.), AI 2003: Advances in Artificial Intelligence. XVI, 1075 pages. 2003.

Vol. 2902: F.M. Pires, S.P. Abreu (Eds.), Progress in Artificial Intelligence. XV, 504 pages. 2003.

Vol. 2892: F. Dau, The Logic System of Concept Graphs with Negation. XI, 213 pages. 2003.

Vol. 2891: J. Lee, M. Barley (Eds.), Intelligent Agents and Multi-Agent Systems. X, 215 pages. 2003.

Vol. 2882: D. Veit, Matchmaking in Electronic Markets. XV, 180 pages. 2003.

Vol. 2871: N. Zhong, Z.W. Raś, S. Tsumoto, E. Suzuki (Eds.), Foundations of Intelligent Systems. XV, 697 pages. 2003.

Vol. 2854: J. Hoffmann, Utilizing Problem Structure in Planing. XIII, 251 pages. 2003.

Vol. 2843: G. Grieser, Y. Tanaka, A. Yamamoto (Eds.), Discovery Science. XII, 504 pages. 2003.

Vol. 2842: R. Gavaldá, K.P. Jantke, E. Takimoto (Eds.), Algorithmic Learning Theory. XI, 313 pages. 2003.

Vol. 2838: N. Lavrač, D. Gamberger, L. Todorovski, H. Blockeel (Eds.), Knowledge Discovery in Databases: PKDD 2003. XVI, 508 pages. 2003.

Vol. 2837: N. Lavrač, D. Gamberger, L. Todorovski, H. Blockeel (Eds.), Machine Learning: ECML 2003. XVI, 504 pages. 2003.

Vol. 2835: T. Horváth, A. Yamamoto (Eds.), Inductive Logic Programming. X, 401 pages. 2003.

Vol. 2821: A. Günter, R. Kruse, B. Neumann (Eds.), KI 2003: Advances in Artificial Intelligence. XII, 662 pages. 2003.

Vol. 2807: V. Matoušek, P. Mautner (Eds.), Text, Speech and Dialogue. XIII, 426 pages. 2003.

Vol. 2801: W. Banzhaf, J. Ziegler, T. Christaller, P. Dittrich, J.T. Kim (Eds.), Advances in Artificial Life. XVI, 905 pages. 2003.

Vol. 2797: O.R. Zaïane, S.J. Simoff, C. Djeraba (Eds.), Mining Multimedia and Complex Data. XII, 281 pages. 2003.

Vol. 2792: T. Rist, R.S. Aylett, D. Ballin, J. Rickel (Eds.), Intelligent Virtual Agents. XV, 364 pages. 2003.

Vol. 2782: M. Klusch, A. Omicini, S. Ossowski, H. Laamanen (Eds.), Cooperative Information Agents VII. XI, 345 pages. 2003.

Vol. 2780: M. Dojat, E. Keravnou, P. Barahona (Eds.), Artificial Intelligence in Medicine. XIII, 388 pages. 2003.

Vol. 2777: B. Schölkopf, M.K. Warmuth (Eds.), Learning Theory and Kernel Machines. XIV, 746 pages. 2003.

Vol. 2752: G.A. Kaminka, P.U. Lima, R. Rojas (Eds.), RoboCup 2002: Robot Soccer World Cup VI. XVI, 498 pages. 2003.

Vol. 2741: F. Baader (Ed.), Automated Deduction – CADE-19. XII, 503 pages. 2003.

Vol. 2705: S. Renals, G. Grefenstette (Eds.), Text- and Speech-Triggered Information Access. VII, 197 pages. 2003.

Vol. 2703: O.R. Zaïane, J. Srivastava, M. Spiliopoulou, B. Masand (Eds.), WEBKDD 2002 - MiningWeb Data for Discovering Usage Patterns and Profiles. IX, 181 pages. 2003.

Vol. 2700: M.T. Pazienza (Ed.), Extraction in the Web Era. XIII, 163 pages. 2003.

Vol. 2699: M.G. Hinchey, J.L. Rash, W.F. Truszkowski, C.A. Rouff, D.F. Gordon-Spears (Eds.), Formal Approaches to Agent-Based Systems. IX, 297 pages. 2002.

Vol. 2691: V. Mařík, J.P. Müller, M. Pechoucek (Eds.), Multi-Agent Systems and Applications III. XIV, 660 pages. 2003.

Vol. 2684: M.V. Butz, O. Sigaud, P. Gérard (Eds.), Anticipatory Behavior in Adaptive Learning Systems. X, 303 pages. 2003.

Vol. 2671: Y. Xiang, B. Chaib-draa (Eds.), Advances in Artificial Intelligence. XIV, 642 pages. 2003.

Vol. 2663: E. Menasalvas, J. Segovia, P.S. Szczepaniak (Eds.), Advances in Web Intelligence. XII, 350 pages. 2003.

Vol. 2661: P.L. Lanzi, W. Stolzmann, S.W. Wilson (Eds.), Learning Classifier Systems. VII, 231 pages. 2003.

Vol. 2654: U. Schmid, Inductive Synthesis of Functional Programs. XXII, 398 pages. 2003.

Vol. 2650: M.-P. Huget (Ed.), Communications in Multiagent Systems. VIII, 323 pages. 2003.

Vol. 2645: M.A. Wimmer (Ed.), Knowledge Management in Electronic Government. XI, 320 pages. 2003.

Vol. 2639: G. Wang, Q. Liu, Y. Yao, A. Skowron (Eds.), Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing. XVII, 741 pages. 2003.

Vol. 2637: K.-Y. Whang, J. Jeon, K. Shim, J. Srivastava, Advances in Knowledge Discovery and Data Mining. XVIII, 610 pages. 2003.

Vol. 2636: E. Alonso, D. Kudenko, D. Kazakov (Eds.), Adaptive Agents and Multi-Agent Systems. XIV, 323 pages. 2003.

Vol. 2627: B. O'Sullivan (Ed.), Recent Advances in Constraints. X, 201 pages. 2003.

Vol. 2600: S. Mendelson, A.J. Smola (Eds.), Advanced Lectures on Machine Learning. IX, 259 pages. 2003.

Vol. 2592: R. Kowalczyk, J.P. Müller, H. Tianfield, R. Unland (Eds.), Agent Technologies, Infrastructures, Tools, and Applications for E-Services. XVII, 371 pages. 2003.

Vol. 2586: M. Klusch, S. Bergamaschi, P. Edwards, P. Petta (Eds.), Intelligent Information Agents. VI, 275 pages. 2003.

Vol. 2583: S. Matwin, C. Sammut (Eds.), Inductive Logic Programming. X, 351 pages. 2003.

Vol. 2581: J.S. Sichman, F. Bousquet, P. Davidsson (Eds.), Multi-Agent-Based Simulation. X, 195 pages. 2003.

Vol. 2577: P. Petta, R. Tolksdorf, F. Zambonelli (Eds.), Engineering Societies in the Agents World III. X, 285 pages. 2003.

Vol. 2569: D. Karagiannis, U. Reimer (Eds.), Practical Aspects of Knowledge Management. XIII, 648 pages. 2002.

Vol. 2560: S. Goronzy, Robust Adaptation to Non-Native Accents in Automatic Speech Recognition. XI, 144 pages. 2002.

Vol. 2557: B. McKay, J. Slaney (Eds.), AI 2002: Advances in Artificial Intelligence. XV, 730 pages. 2002.

Vol. 2554: M. Beetz, Plan-Based Control of Robotic Agents. XI, 191 pages. 2002.

Vol. 2543: O. Bartenstein, U. Geske, M. Hannebauer, O. Yoshie (Eds.), Web Knowledge Management and Decision Support. X, 307 pages. 2003.

Vol. 2541: T. Barkowsky, Mental Representation and Processing of Geographic Knowledge. X, 174 pages. 2002.

Vol. 2533: N. Cesa-Bianchi, M. Numao, R. Reischuk (Eds.), Algorithmic Learning Theory. XI, 415 pages. 2002.

Vol. 2531: J. Padget, O. Shehory, D. Parkes, N.M. Sadeh, W.E. Walsh (Eds.), Agent-Mediated Electronic Commerce IV. Designing Mechanisms and Systems. XVII, 341 pages. 2002.

Vol. 2527: F.J. Garijo, J.-C. Riquelme, M. Toro (Eds.), Advances in Artificial Intelligence - IBERAMIA 2002. XVIII, 955 pages. 2002.

Vol. 2522: T. Andreasen, A. Motro, H. Christiansen, H.L. Larsen (Eds.), Flexible Query Answering Systems. X, 383 pages. 2002.

Vol. 2514: M. Baaz, A. Voronkov (Eds.), Logic for Programming, Artificial Intelligence, and Reasoning. XIII, 465 pages. 2002.

Vol. 2507: G. Bittencourt, G.L. Ramalho (Eds.), Advances in Artificial Intelligence. XIII, 417 pages. 2002.

Vol. 2504: M.T. Escrig, F. Toledo, E. Golobardes (Eds.), Topics in Artificial Intelligence. XI, 427 pages. 2002.

Vol. 2499: S.D. Richardson (Ed.), Machine Translation: From Research to Real Users. XXI, 254 pages. 2002.

Vol. 2484: P. Adriaans, H. Fernau, M. van Zaanen (Eds.), Grammatical Inference: Algorithms and Applications. IX, 315 pages. 2002.

Preface

Human conversational partners are able, at least to a certain extent, to detect the speaker's or listener's emotional state and may attempt to respond to it accordingly. When instead one of the interlocutors is a computer a number of questions arise, such as the following: To what extent are dialogue systems able to simulate such behaviors? Can we learn the mechanisms of emotional behaviors from observing and analyzing the behavior of human speakers? How can emotions be automatically recognized from a user's mimics, gestures and speech? What possibilities does a dialogue system have to express emotions itself? And, very importantly, would emotional system behavior be desirable at all?

Given the state of ongoing research into incorporating emotions in dialogue systems we found it timely to organize a Tutorial and Research Workshop on Affective Dialogue Systems (ADS 2004) at Kloster Irsee in Germany during June 14–16, 2004. After two successful ISCA Tutorial and Research Workshops on Multimodal Dialogue Systems at the same location in 1999 and 2002, we felt that a workshop focusing on the role of affect in dialogue would be a valuable continuation of the workshop series.

Due to its interdisciplinary nature, the workshop attracted submissions from researchers with very different backgrounds and from many different research areas, working on, for example, dialogue processing, speech recognition, speech synthesis, embodied conversational agents, computer graphics, animation, user modelling, tutoring systems, cognitive systems, and human-computer interaction. Overall, ADS 2004 embodied 23 long papers, and 12 short and demonstration papers, not only from 10 Western and Eastern European countries, but also from Canada, Japan and the US. The papers cover the following seven topic areas:

- emotion recognition
- affective user modelling
- affective conversational agents and dialogue simulation
- emotional databases, annotation schemes and tools
- synthesis of emotional speech and facial animation
- affective tutoring systems
- evaluation of affective dialogue systems

The main characteristic of the papers in this volume is their endeavor to go beyond pure task-oriented approaches to dialogue processing to also address conversational aspects as well as psychological and social concerns. For example the volume includes a number of empirical studies that investigate in what sense the human user could benefit from the imitation of social behaviors.

A theme that came up in a number of contributions in this volume is the important role of emotional factors in the development of embodied conversational agents. Many authors are convinced that the integration of an affective component can significantly enhance an agent's believability.

For an agent to behave human-like, appropriate synchronization of speech, mimics and gestures is highly important and is a major concern in research on affective dialogue systems. But also research on the individual output components is crucial to the advances in such systems. In this volume this is witnessed by a number of papers, not least on emotional speech synthesis.

The work on tutoring systems presented in this volume indicates that agents that attend to and attempt to influence the motivational state of students may aid the learning process.

In addition to work on the expression of emotions, papers on the recognition and interpretation of the user's emotional state are presented in this volume. Interestingly, the authors approach this topic from various perspectives. One group focuses on the signal-processing level, aiming to recognize typical emotional patterns from biosensors, facial expressions and speech. Another group extends traditional work on user modelling with an affective component.

Finally, the volume includes papers on the evaluation of affective dialogue systems. Not least on the usability side, this is an area with many open ends.

We would like to thank all authors for the effort they spent on their submissions, and the program committee – more than 30 distinguished researchers from industry and academia – who worked very hard to tight deadlines and selected the best contributions for the final program. We are also grateful to the invited speakers for enriching our workshop.

In addition, we would like to express our thanks to several people who assisted us in organizing the workshop. Torben Kruchov Madsen took care of the Web page for uploading papers. Angela Rittinger and Brigitte Waimer-Eichenauer provided worthwhile administrative support. A number of organizations supported ADS 2004 including ACL Sigmedia, ACL/ISCA Sigdial, Gesellschaft für Informatik (GI), and ISCA. In particular, we gratefully acknowledge GI for their valuable assistance in handling the financial matters. Last, but not least, we are grateful to Springer-Verlag for publishing the proceedings in their LNCS/LNAI series.

April 2004

Elisabeth André
Laila Dybkjær
Paul Heisterkamp
Wolfgang Minker

Committee Listings

Organizing Committee

Elisabeth André (University of Augsburg, Germany)
Laila Dybkjær (University of Southern Denmark, Denmark)
Paul Heisterkamp (DaimlerChrysler Research and Technology, Germany)
Wolfgang Minker (University of Ulm, Germany)

Program Committee

Jan Alexandersson
Niels Ole Bernsen
Dirk Bühler
Cristina Conati
Janet Cahn
Piero Cosi
Klaus Dorfmüller-Ulhaas
Ellen Douglas-Cowie
Silke Goronzy
Jim Glass
Björn Granström
Jon Gratch
Joakim Gustafson
Eli Hagen
Lisa Harper
Kristina Höök
David House
Ralf Kompe
Christine Lisetti
Jean-Claude Martin
Dominic Massaro
Géza Németh
Elmar Nöth
Ana Paiva
Catherine Pelachaud
Helmut Prendinger
Fiorella de Rosis
Alex Rudnicky
Marc Schröder
Oliviero Stock
David Traum
Wolfgang Wahlster
Wayne Ward

Sponsoring Institutions

Gesellschaft für Informatik (GI)
ACL and ISCA Special Interest Group on Discourse and Dialogue (SIGDIAL)
ACL Special Interest Group on Multimedia Language Processing (SIGMEDIA)
International Speech Communication Association (ISCA)
DaimlerChrysler Research and Technology, Dialogue Systems
University of Augsburg, Dept. of Multimedia Concepts and Applications
University of Southern Denmark, Natural Interactive Systems Laboratory
University of Ulm, Department of Information Technology

Table of Contents

Emotion Recognition

Affective User Modeling

Emotional Databases, Annotation Schemes, and Tools

Affective Conversational Agents and Dialogue Simulation

Synthesis of Emotional Speech and Facial Animations

Affective Tutoring Systems

Evaluation of Affective Dialogue Systems

Demonstrations

From Emotion to Interaction: Lessons from Real Human-Machine-Dialogues

Anton Batliner[1], Christian Hacker[1], Stefan Steidl[1], Elmar Nöth[1], and Jürgen Haas[2]

[1] University of Erlangen-Nuremberg, Lehrstuhl für Mustererkennung / Chair for Pattern Recognition, Martensstr.3, 91058 Erlangen, F.R.G
batliner@informatik.uni-erlangen.de
[2] Sympalog Voice Solutions GmbH, Karl Zuckerstr. 10, Erlangen, F.R.G.

Abstract. The monitoring of emotional user states can help to assess the progress of human-machine-communication. If we look at specific databases, however, we are faced with several problems: users behave differently, even within one and the same setting, and some phenomena are sparse; thus it is not possible to model and classify them reliably. We exemplify these difficulties on the basis of SympaFly, a database with dialogues between users and a fully automatic speech dialogue telephone system for flight reservation and booking, and discuss possible remedies.

1 Introduction[1]

It might be fair to describe one (maybe 'the') basic conceptualization of using information on emotions within automatic dialogue systems in the following way: if we detect something like anger, let's initiate some recovery strategy or hand over to a human operator. If we detect something like joy, try to utilize this information, for instance, by offering some new, good bargain. This seems to be a realistic vision if we consider the good classification rates obtained for some basic emotions in the laboratory. As far as we can see, however, the few studies conducted during the last years dealing with non-acted emotions recorded in a realistic scenario report rather a negative correlation between full-blown, prototypical emotions on the one hand, and frequency on the other hand; moreover, the recognition rates for real-life speech data go down considerably, cf. [2,5,1,9]. We believe, that a way out of this dilemma is not only to collect more data but first of all, to take into account more phenomena: the monitoring of the user's behavior should not only consider some basic emotions but all kind of emotional user states, and in addition, we should look for any change in the user's behavior towards other 'suspicious' directions, e.g., use of meta-talk or of repetitions.

[1] This work was funded by the EU in the project PF-STAR (*http://pfstar.itc.it/*) under grant IST-2001-37599 and by the German Federal Ministry of Education, Science, Research and Technology (*BMBF*) in the SmartKom project under Grant 01 IL 905 K7. The responsibility for the contents of this study lies with the authors.

E. André et al. (Eds.): ADS 2004, LNAI 3068, pp. 1–12, 2004.

Thus, the focus of interest has to be shifted from a subject-centered towards an interaction-centered point of view, cf. section 6.

In this paper which is reporting work in progress, we first present SympaFly, a fully automatic speech dialogue telephone system for flight reservation and booking. In the first stage of this system, performance was rather poor (approx. 30% dialogue success rate); in the last stage, performance was very good (above 90% dialogue success rate). All dialogues were orthographically transliterated and annotated as for (emotional) user states, prosodic peculiarities, dialogue (step) success rate, and conversational peculiarities. For classification of user states, a large prosodic feature vector was used. We will show that users employ different strategies, and that it is really mandatory to deal with the sparse data problem as far as emotional user states are concerned.

2 The SympaFly Database

SympaFly is a fully automatic speech dialogue telephone system for flight reservation and booking. The database comprises three different stages; the methodology consisted of a rapid prototyping phase followed by optimization iterations. Subjects were asked to call the automatic dialogue system and book one or more flights. The caller should, for instance, book a flight from Zurich to Tiflis and back so that the meeting there can take place at a specific time. Additional information had to be given, e.g., frequent flyer id, credit card number, and so on. The three evaluation stages can be characterized as follows; a more detailed account of the system design can be found in [4]:

- The first part of the data set **S1** (110 dialogues, 2291 user turns, 11581 words; 5.1 words per turn, 105 words and 20.8 turns per dialogue) are those dialogues which were collected in the first test of the system, conducted by an independent usability lab, built by only using the input of involved system developers and designers, without any external evaluation whatsoever. The performance of the system was rather poor.
- The dialogues in the second phase **S2** (annotated and processed: 98 dialogues, 2674 user turns, 9964 words; 3.7 words per turn, 102 words and 27.3 turns per dialogue) cover several system phases, wherein the system performance was increased little by little, sometimes from one day to the other. Due to this, the individual dialogues can strongly differ depending on the system performance at a particular time. Callers were volunteers without any connection with the usability lab.
- Finally, the third part **S3** (62 dialogues, 1900 user turns, 7655 words; 4.0 words per turn, 124 words and 30.6 turns per dialogue) contains dialogues collected through the final system, by using the same experimental setting as for S1: same telephone channel, callers are supervised by the usability lab. The performance of the system was now excellent.

3 Annotations and Feature Extraction

For the annotation of **holistic (emotional) user states**, no pre-defined set of labels was given; two labellers decided themselves which and how many different user states to annotate; interlabeller correspondence is discussed in [4]. After a first, independent run the labellers decided on a consensus labelling in a second run. The following turn-based labels (given in italics) were used and mapped onto these five cover classes (given recte and in boldface): **positive:** ***Joyful***; **neutral:** ***Neutral***; **pronounced:** ***Emphatic***; **weak negative:** ***Surprised, Ironic,*** **strong negative:** ***Helpless, Panic, Touchy*** (i.e., irritated), ***Angry. Emphatic*** is taken as sort of 'basically suspicious' – in our scenario most likely not positive, but indicating problems; this assumption will be discussed further below.

It can be assumed that users encounting difficulties in the communication with a system, change their way of speaking, for instance, by emphasising salient information. In Table 1[2], the labels used for the annotation of such **prosodic peculiarities** are given, arranged according to their presumed strength; labels covering more than one strength level can be either the one or the other level. (For a two-class problem, the three labels given in italics could be attributed to the (cover) class *neutral*.) Laughter and syllable lengthening cannot be attributed to one specific level of prosodic strength. More than one label can be attributed to the same word; in such a case, for the mapping onto strength levels, the strongest one 'wins'. This is again a consensus labelling of two annotators. The label set has been used in the Verbmobil- and in the SmartKom-project [2,11].

Table 1. Prosodic peculiarities, annotated word-based, and their strength

weak	medium	strong
pause_phrase	pause_word	pause_syllable
emphasis		strong emphasis
clear_articulation		hyper-articulation
lengthening_syllable		
laughter		

Another labeller annotated the **success of a dialogue** using four levels: *null* (no user confirmation, no booking), *full* (confirmation and booking), and two levels in between: *some* (maybe confirmation but no booking), and *medium* (confirmation, but no 'ideal' booking). In addition to this global measure, we annotate for each turn ten slots that can - but need not - be filled in each user utterance: *departure*, *destination*, *date*, *time*, *class*, *persons*, *membership* (in the frequent flyer program), *number of membership*, *credit-card number*, *credit-card validity*. These slot fillers can be compared with the preceding system utterance, and then we can decide whether a dialogue step has been successful or not.

[2] 'pause_phrase': extra long pause between syntactic units, 'pause_word': pause between words inside syntactic unit; 'pause_syll': pause inside word; the other labels are self-explanatory.