

## Appendix A.

1	Welcome to the CMU Let's Go bus information system.	18	Leaving from DYNAMO. Is this correct?
2	To get help at any time, just say Help or press zero.		<i>No.</i>
3	What can I do for you?	19	Leaving from DYNAMO. Is this correct?
	<DTMF-Zero>		<i>Leaving to airport.</i>
4	I am an automated spoken dialogue system that can give you schedule information for bus routes in Pittsburgh's East End. You can ask me about the following buses: 28X, 54C, 56U, 59U, 61A, 61B, 61C, 61D, 61F, 64A, 69A, and 501.	20	Leaving from AIRPORT . Did I get that right?
5	What bus schedule information are you looking for?		<i>yes</i>
6	For example, you can say, when is the next 28X from DOWNTOWN to THE AIRPORT? or I'd like to go from MCKEESPORT to HOME-STEAD tomorrow at 10 a.m..	21	Right
	28X	22	Where are you going?
7	The 56C . Is this correct?		<i>To airport.</i>
	<i>No, 28X</i>	23	Going to AIRPORT. Did I get that right?
8	What can I do for you?		<i>yeah, yep</i>
	<i>I want to know whether /uh/ tomorrow the bus is available or not.</i>	24	Okay
9	Leaving at 11 p.m. . Did I get that right?	25	When would you like to travel?
	<i>No, I just want to know is there any bus service for 28X tomorrow.</i>		<i>yes</i>
10	Please use shorter answers because I have trouble understanding long sentences.	26	When would you like to travel?
	Leaving at 11 p.m. . Is this correct?		<i>Twenty third November.</i>
	<i>On twenty third november.</i>	27	When would you like to travel?
12	Leaving at 11 p.m. . Is this correct?		<i>Tomorrow.</i>
	<i>No, I want to know the schedule on twenty third november</i>	28	At what time do you wanna travel?
13	Leaving at 2 a.m. . Did I get that right?		<i>Seven a.m.</i>
	<noise>	29	Leaving at 7 a.m. . Is this correct?
14	If you want to leave at 2 a.m. say yes or press one, otherwise say no or press three.		<i>yep</i>
	<i>no &lt;noise&gt;</i>	30	Alright
15	Going to NORTH AND CEDAR. Did I get that right?	31	Just a minute. I'll look that up.
	<i>28x going to airport at seven a.m.</i>	32	AIRPORT and AIRPORT are both the same stop. Please provide a different start or end point.
16	Please use shorter answers because I have trouble understanding long sentences.	33	To ask about a different trip, you can say, start a new query. If you are finished, you can say goodbye.
17	Going to NORTH AND CEDAR. Is this correct?		<i>Yes.</i>
	<i>No.</i>	34	You can say start a new query, or goodbye.
			<i>No.</i>
		35	To ask about a different trip, you can say, start a new query. If you are finished, you can say goodbye.
			<noise>
		36	You can say start a new query, or goodbye.
			<noise>
			You can say start a new query, or goodbye.
			<hangup>

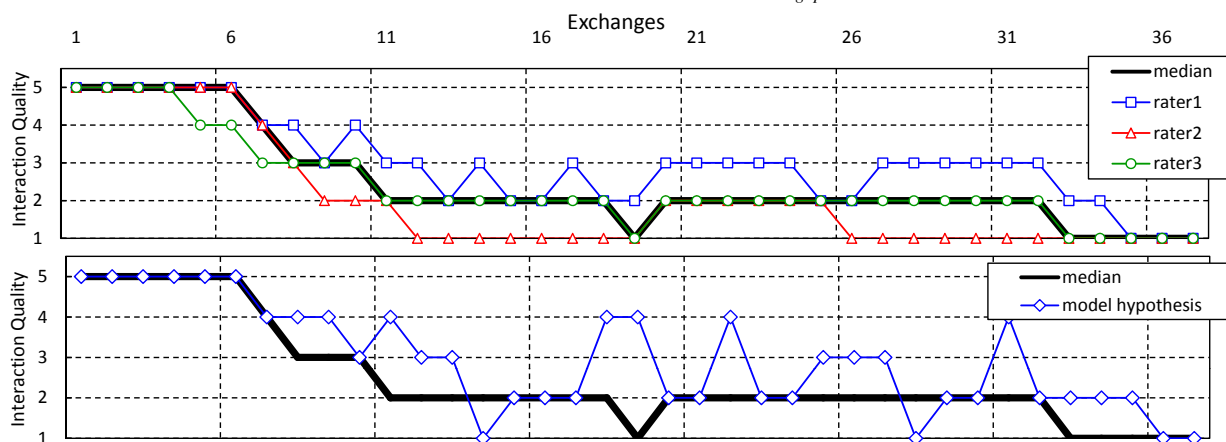


Table A.1: Example dialogue (ID: 2061122025) from the CMU Let's Go System (2006 corpus) with low Interaction Quality. The user utterances are printed in *italics*. Upper chart: Turn-wise Interaction Quality (IQ) annotation from 3 raters. The final label is the median of all three opinions. Lower chart: median reference vs. hypothesis of the model trained with *AUTO* feature set.

## Appendix B.

Parameter	Description
SEMANTICPARSE	Semantic parse of the caller utterance as returned by the activated grammar. EL: Litman et al. (1999)
HELPREQUEST?	Current turn is a (from the system recognized) help request, i.e. the user asks for more details ( $\in \{true, false\}$ ).
#HELP REQUESTS	Number of help requests up to this exchange $e_n$ . $\sum_{i=1}^n x \begin{cases} 1 & \text{HELPREQUEST?}=true, \\ 0 & \text{otherwise.} \end{cases}$ EL: Paek and Horvitz (2004) DL: Kamm et al. (1998); Litman and Pan (1999); International Telecommunication Union (2005); Hajdinjak and Mihelic (2006)
{#}HELPREQUESTS	Number of help requests within previous $w$ turns prior to $e_n$ , where $w$ is the size of the window. $\sum_{i=n-w}^n x \begin{cases} 1 & \text{HELPREQUEST?}=true, \\ 0 & \text{otherwise.} \end{cases}$
% HELP REQUESTS	Percentage of help-requests in all previous exchanges: $\frac{1}{n} \sum_{i=1}^n x \begin{cases} 1 & \text{HELPREQUEST?}=true \\ 0 & \text{otherwise.} \end{cases}$ EL: Paek and Horvitz (2004) DL: Litman et al. (1999); Hajdinjak and Mihelic (2006)
OPERATOR REQUEST?	Current turn is a (from the system recognized) request for an operator, i.e. the user opts out ( $\in \{true, false\}$ ). EL: Paek and Horvitz (2004)
# OPERATOR REQUESTS	Number of operator requests up to this exchange $e_n$ . $\sum_{i=1}^n x \begin{cases} 1 & \text{OPERATORREQUEST?}=true, \\ 0 & \text{otherwise.} \end{cases}$ EL: Paek and Horvitz (2004)

Table B.1: (cont.) Interaction parameters derived from the LU module for employment on exchange level.

Parameter	Description
$\{\#\}_{\text{OPERATOR REQUESTS}}$	<p>Number of operator requests within previous <math>w</math> turns prior to <math>e_n</math>, where <math>w</math> is the size of the window.</p> $\sum_{i=n-w}^n x \begin{cases} 1 & \text{OPERATORREQUEST?}=true, \\ 0 & \text{otherwise.} \end{cases}$
$\% \text{ OPERATOR REQUESTS}$	<p>Percentage of operator requests in all previous exchanges:</p> $\frac{1}{n} \sum_{i=1}^n x \begin{cases} 1 & \text{OPERATOR?}=true \\ 0 & \text{otherwise.} \end{cases}$ <p>EL: Paek and Horvitz (2004)</p>

Table B.1: Parameters derived from logs from the LU module.

Parameter	Description
GRAMMAR	<p>Names of all activated grammars.</p> <p>EL: Langkilde et al. (1999); Walker et al. (2002) DL: Litman et al. (2000)</p>
TRIGGERED GRAMMAR	<p>Name of grammar that returned the ASR parse.</p> <p>EL: Langkilde et al. (1999); Walker et al. (2002) DL: Riccardi et al. (2000)</p>
UTTERANCE	<p>ASR parse from user utterance, i.e. automatically transcribed.</p> <p>EL: Langkilde et al. (1999); Walker et al. (2002); Levin and Pieraccini (2006)</p>
ASRRECOGNITIONSTATUS	<p>Status of the ASR when trying to parse the user input <math>\in \{success, reject, timeout\}</math> “success” refers to a - from the ASR’s point-of-view - successful recognition of the utterance, i.e. the decoded word string matches an active grammar, “reject” means the ASR could not recognize the utterance and did not find a corresponding word sequence according to the active grammars; “timeout” indicates that the user did not respond within a given time slot.</p> <p>EL: Langkilde et al. (1999); Walker et al. (2002); Levin and Pieraccini (2006)</p>

Table B.2: (cont.) Interaction parameters derived from the ASR module for employment on exchange level.

Parameter	Description
#ASRSUCCESS	<p>Number of successfully parsed turns up to this exchange <math>e_n</math>.</p> $\sum_{i=1}^n x \begin{cases} 1 & \text{ASRRECOGNITIONSTATUS}=\text{success}, \\ 0 & \text{otherwise.} \end{cases}$ <p>EL: Paek and Horvitz (2004) DL: International Telecommunication Union (2005)</p>
{#}ASRSUCCESS	<p>Number of successfully parsed turns within previous <math>w</math> turns prior to <math>e_n</math>, where <math>w</math> is the size of the window.</p> $\sum_{i=n-w}^n x \begin{cases} 1 & \text{ASRRECOGNITIONSTATUS}=\text{success} \\ 0 & \text{otherwise.} \end{cases}$
%ASRSUCCESS	<p>Percentage of “Success” turns in all previous exchanges:</p> $\frac{1}{n} \sum_{i=1}^n x \begin{cases} 1 & \text{ASRRECOGNITIONSTATUS}=\text{success} \vee \\ & \text{ACTIVITYTYPE} \notin \{\text{question}, \text{confirmation}\} \\ 0 & \text{otherwise.} \end{cases}$ <p>EL: Paek and Horvitz (2004) DL: Litman et al. (1999); Litman and Pan (2002); International Telecommunication Union (2005)</p>
#TIMEOUTPROMPTS	<p>Number of time-out events up to this exchange <math>e_n</math>.</p> $\sum_{i=1}^n x \begin{cases} 1 & \text{ASRRECOGNITIONSTATUS}=\text{time-out}, \\ 0 & \text{otherwise.} \end{cases}$ <p>EL: Paek and Horvitz (2004); Kim (2007) DL: Kamm et al. (1998); Litman and Pan (1999); International Telecommunication Union (2005)</p>
{#}TIMEOUTPROMPTS	<p>Number of time-out turns within previous <math>w</math> turns prior to <math>e_n</math>, where <math>w</math> is the size of the window.</p> $\sum_{i=n-w}^n x \begin{cases} 1 & \text{ASRRECOGNITIONSTATUS}=\text{timeout}, \\ 0 & \text{otherwise.} \end{cases}$

Table B.2: (cont.) Interaction parameters derived from the ASR module for employment on exchange level.

Parameter	Description
%TIMEOUTPROMPTS	<p>Percentage of “time-out” turns in all previous exchanges:</p> $\frac{1}{n} \sum_{i=1}^n x \begin{cases} 1 & \text{ASRRECOGNITIONSTATUS=time-out} \\ 0 & \text{otherwise.} \end{cases}$ <p><math>x = 0 \forall</math> exchanges in the beginning of the dialogue where no ASR is active.</p> <p>EL: Paek and Horvitz (2004) DL: Litman et al. (1999); Walker et al. (2000)</p>
#ASRREJECTIONS	<p>Number of ASR rejections up to this exchange <math>e_n</math>.</p> $\sum_{i=1}^n x \begin{cases} 1 & \text{ASRRECOGNITIONSTATUS=reject,} \\ 0 & \text{otherwise.} \end{cases}$ <p>EL: Paek and Horvitz (2004); Levin and Pieraccini (2006) DL: Kamm et al. (1998); Litman et al. (1999); Litman and Pan (1999); Walker et al. (2000); International Telecommunication Union (2005)</p>
{#}ASRREJECTIONS	<p>Number of ASR rejections within the previous <math>w</math> turns prior to <math>e_n</math>, where <math>w</math> is the size of the window.</p> $\sum_{i=n-w}^n x \begin{cases} 1 & \text{ASRRECOGNITIONSTATUS=reject,} \\ 0 & \text{otherwise.} \end{cases}$
#TIMEOUTS_ASRREJ	<p>Number of time-out and ASR rejection events up to this exchange <math>e_n</math>.</p> $\sum_{i=1}^n x \begin{cases} 1 & \text{ASRRECOGNITIONSTATUS} \in \{timeout, reject\}, \\ 0 & \text{otherwise.} \end{cases}$
{#}TIMEOUTS_ASRREJ	<p>Number of time-out and ASR rejection events within the previous <math>w</math> turns prior to <math>e_n</math>, where <math>w</math> is the size of the window.</p> $\sum_{i=n-w}^n x \begin{cases} 1 & \text{ASRRECOGNITIONSTATUS} \in \{timeout, reject\}, \\ 0 & \text{otherwise.} \end{cases}$
% TIMEOUTS _ASRREJ	<p>Percentage of time-out and ASR rejection events in all previous exchanges:</p> $\frac{1}{n} \sum_{i=1}^n x \begin{cases} 1 & \text{ASRRECOGNITIONSTATUS} \in \{timeout, reject\} \\ 0 & \text{otherwise.} \end{cases}$ <p>Exchanges in the beginning of the dialogue where no ASR is active are replenished with “0”.</p>

Table B.2: (cont.) Interaction parameters derived from the ASR module for employment on exchange level.

Parameter	Description
BARGE-IN?	True if the user interrupted the system prompt, false otherwise ( $\in \{true, false\}$ )
# BARGE-INS	<p>Number of barge-ins up to this exchange <math>e_n</math>.</p> $\sum_{i=1}^n x \begin{cases} 1 & \text{BARGEIN?}=true, \\ 0 & \text{BARGEIN?}=false. \end{cases}$ <p>DL: Kamm et al. (1998); Litman et al. (1999); Litman and Pan (1999); Walker et al. (2000); International Telecommunication Union (2005)</p>
{#} BARGE-INS	<p>Number of barge-ins within previous <math>w</math> turns prior to <math>e_n</math>, where <math>w</math> is the size of the window.</p> $\sum_{i=n-w}^n x \begin{cases} 1 & \text{BARGEIN?}=true, \\ 0 & \text{BARGEIN?}=false. \end{cases}$
% BARGE-INS	<p>Percentage of barge-ins in all previous exchanges:</p> $\frac{1}{n} \sum_{i=1}^n x \begin{cases} 1 & \text{BARGEIN?}=true, \\ 0 & \text{BARGEIN?}=false. \end{cases}$ <p>DL: Litman et al. (1999); Walker et al. (2000)</p>
ASRCONFIDENCE	<p>Confidence of the ASR module representing the certainty of returning the correct ASR parse (<math>\in \mathbb{R}\{0..1\}</math>).</p> <p>EL: Langkilde et al. (1999); Paek and Horvitz (2004); Levin and Pieraccini (2006) DL: Kamm et al. (1998); Walker et al. (2000); Litman and Pan (2002)</p>
MEANASRCONFIDENCE	<p>Average ASR confidence up to this exchange.</p> $\frac{1}{n} \sum_{i=1}^n x \begin{cases} ASRConfidence_i & ASRRECOGNITIONSTATUS \in \{success, reject\} \\ ASRConfidence_{corpus\_mean} & \text{otherwise.} \end{cases}$ <p>Missing values are replenished with the average corpus-wide confidence <math>ASRConfidence_{corpus\_mean}</math> calculated on all “Success” and “Reject” turns.</p> <p>DL: Litman et al. (1999); Litman and Pan (1999)</p>

Table B.2: (cont.) Interaction parameters derived from the ASR module for employment on exchange level.

Parameter	Description
{MEAN}ASRCONFIDENCE	<p>Average ASR confidence within previous <math>w</math> turns prior to <math>e_n</math>, where <math>w</math> is the size of the window.</p> $\frac{1}{n} \sum_{i=n-w}^n x \begin{cases} ASRConfidence_i & ASRREC.STATUS \in \{success, reject\} \\ ASRConfidence_{corpus\_mean} & \text{otherwise.} \end{cases}$ <p>Missing values are replenished with the average corpus-wide confidence <math>ASRConfidence_{corpus\_mean}</math> calculated on all “Success” and “Reject” turns.</p>
UTD	<p>Utterance Turn Duration of the user utterance in seconds.</p> <p>EL: Langkilde et al. (1999); Walker et al. (2002); Levin and Pieraccini (2006) DL: International Telecommunication Union (2005)</p>
ExMo	<p>Expected input modality by the system at current exchange <math>\in \{speech, dtmf, both, none\}</math>.</p> <p>EL: Langkilde et al. (1999); Walker et al. (2002)</p>
MODALITY	<p>Input modality of the user for responding to a question <math>\in \{speech, dtmf\}</math>.</p> <p>EL: Langkilde et al. (1999); Walker et al. (2002); Levin and Pieraccini (2006)</p>
UNExMo?	<p>User employed other modality than suggested by system prompt.</p> $\begin{cases} true & MODALITY \notin EXMO \\ false & \text{otherwise.} \end{cases}$
#UNExMo	<p>Number of unexpected modality usage up to this exchange <math>e_n</math>.</p> $\sum_{i=1}^n x \begin{cases} 1 & MODALITY \notin EXMO \\ 0 & \text{otherwise.} \end{cases}$
{#}UNExMo	<p>Number of unexpected modality usages within previous <math>w</math> turns prior to <math>e_n</math>, where <math>w</math> is the size of the window.</p> $\sum_{i=n-w}^n x \begin{cases} 1 & MODALITY \notin EXMO \\ 0 & \text{otherwise.} \end{cases}$

Table B.2: (cont.) Interaction parameters derived from the ASR module for employment on exchange level.

Parameter	Description
%UNExMo	Percentage of unexpected modality usage up to this exchange. $\frac{1}{n} \sum_{i=1}^n x \begin{cases} 1 & \text{MODALITY} \notin \text{ExMo} \\ 0 & \text{otherwise.} \end{cases}$
WPUT	Number of words returned in the parse. EL: Langkilde et al. (1999); Walker et al. (2002); Levin and Pieraccini (2006) DL: International Telecommunication Union (2005)

Table B.2: Interaction parameters derived from the ASR module for employment on exchange level.

Parameter	Description
ACTIVITY	The name of the activity that was performed by the system consisting of an identifier for the question or statement. Activities of different dialogue systems are according to the flow chart design respectively, e.g. a bus information service may have activities like “query.arrival_place”, “query.travel_time”, “confirm.okay” etc. The names are determined by the system designer. EL: Langkilde et al. (1999); Walker et al. (2002)
ACTIVITYTRIGRAM	Sequence of the current activity plus the two previous activities. This feature models the history of activities.
ACTIVITYTYPE	Type of activities $\in \{\text{“announcement”}, \text{“question”}, \text{“confirmation”}, \text{“wait”}\}$ .
DD	Dialogue duration in seconds up to this exchange. EL: Langkilde et al. (1999); Walker et al. (2002) DL: Walker et al. (2000); International Telecommunication Union (2005)
PROMPT	System prompt of the automated agent prior to recording user input. EL: Langkilde et al. (1999); Walker et al. (2002)
ROLE INDEX	In a dialog module activity, the number of tries to elicit a desired response from the user.
ROLE NAME	The function of this system turn $\in \{\text{“collection”}, \text{“confirmation”}\}$ .

Table B.3: (cont.) Interaction parameters derived from the DM module for employment on exchange level.



Parameter	Description
REPROMPT?	Current turn is a <i>reprompt</i> $\in \{true, false\}$ . EL: Langkilde et al. (1999); Walker et al. (2002)
# REPROMPT	Number of reprompts up to this exchange $e_n$ . $\sum_{i=1}^n x \begin{cases} 1 & \text{REPROMPT?}=true, \\ 0 & \text{otherwise.} \end{cases}$ EL: Langkilde et al. (1999); Walker et al. (2002); Levin and Pieraccini (2006)
{#}REPROMPT	Number of reprompts within previous $w$ turns prior to $e_n$ , where $w$ is the size of the window. $\sum_{i=n-w}^n x \begin{cases} 1 & \text{REPROMPT?}=true, \\ 0 & \text{otherwise.} \end{cases}$
% REPROMPT	Percentage of reprompts in all previous exchanges: $\frac{1}{n} \sum_{i=1}^n x \begin{cases} 1 & \text{REPROMPT?}=true \\ 0 & \text{otherwise.} \end{cases}$ EL: Langkilde et al. (1999); Walker et al. (2002)
LOOP NAME	Given caller speech input, we need to try and recognize the semantic meaning. The first time of try is indicated with a value of 'Initial'. If the system could not turn a parse then we have to re-prompt ("Retry1" or "Timeout1"). Similar for if the caller asks for help or a repetition of the prompt.
EXCHANGES, #SYSTEMTURNS, #USERTURNS	Number of exchanges and system/user turns up to this exchange. The number of turns may differ from the number of exchanges since exchanges may also exist of a single user or system turn. EL: Langkilde et al. (1999); Walker et al. (2002); Levin and Pieraccini (2006) DL: Litman and Pan (1999); International Telecommunication Union (2005); Möller et al. (2008)
# SYSTEM QUESTIONS	Number of system questions up to this exchange. DL: International Telecommunication Union (2005)
{#} SYSTEM QUESTIONS	Number of system questions in the current sliding window.

Table B.3: (cont.) Interaction parameters derived from the DM module for employment on exchange level.

Parameter	Description
WPST	Words per system turn. EL: Litman et al. (1999) DL: International Telecommunication Union (2005)

Table B.3: Parameters collected in DM module used in online prediction

- , 1994. Terms and Definitions related to Quality of Service and Network Performance Including Dependability. ITU-T Recommendation E.800. International Telecommunication Union. Geneva, Switzerland.
- , 2007. Vocabulary for performance and quality of service. ITU-T Amendment Amendment 1 to P.10/G.100. International Telecommunication Union. Geneva, Switzerland.
- Bennett, K.P., Campbell, C., 2000. Support vector machines: Hype or hallelujah? *Journal of SIGKDD Explorations* 2, 1–13.
- Black, A., Eskenazi, M., 2009. The spoken dialogue challenge, in: *Proceedings of the SIGDIAL 2009 Conference*, Association for Computational Linguistics, London, UK. pp. 337–340.
- Carletta, J., 1996. Assessing agreement on classification tasks: the kappa statistic. *Comput. Linguist.* 22, 249–254.
- Cohen, P.R., 1995. *Empirical methods for artificial intelligence*. MIT Press, Cambridge, MA, USA.
- Danieli, M., Gerbino, E., 1995. Metrics for evaluating dialogue strategies in a spoken language system, in: *Working Notes AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, pp. 34–39.
- Doll, W.J., Torkzadeh, G., 1991. The measurement of end-user computing satisfaction: theoretical and methodological issues. *MIS Q.* 15, 5–10.
- Engelbrecht, K.P., Gödde, F., Hartard, F., Ketabdar, H., Möller, S., 2009. Modeling user satisfaction with hidden markov model, in: *SIGDIAL '09: Proceedings of the SIGDIAL 2009 Conference*, Association for Computational Linguistics, Morristown, NJ, USA. pp. 170–177.
- Engelbrecht, K.P., Kühnel, C., Möller, S., 2008. Weighting the coefficients in paradise models to increase their generalizability, in: *Proceedings of the 4th IEEE tutorial and research workshop on Perception and Interactive Technologies for Speech-Based Systems: Perception in Multimodal Dialogue Systems*, Springer-Verlag, Berlin, Heidelberg. pp. 289–292.
- Evani, K., Hunter, P., Liscombe, J., Sündermann, D., Dayanidhi, K., Pieraccini, R., 2008. Caller experience: A method for evaluating dialog systems and its automatic prediction.
- Hajdinjak, M., Mihelic, F., 2006. The paradise evaluation framework: Issues and findings. *Comput. Linguist.* 32, 263–272.
- Hara, S., Kitaoka, N., Takeda, K., 2010. Estimation method of user satisfaction using n-gram-based dialog history model for spoken dialog system, in: *Chair*, N.C.C., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., Tapias, D. (Eds.), *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, European Language Resources Association (ELRA), Valletta, Malta.
- Higashinaka, R., Minami, Y., Dohsaka, K., Meguro, T., 2010a. Issues in predicting user satisfaction transitions in dialogues: Individual differences, evaluation criteria, and prediction models, in: Lee, G., Mariani, J., Minker, W., Nakamura, S. (Eds.), *Spoken Dialogue Systems for Ambient Environments*. Springer Berlin / Heidelberg, volume 6392 of *Lecture Notes in Computer Science*, pp. 48–60. 10.1007/978-3-642-16202-2.5.
- Higashinaka, R., Minami, Y., Dohsaka, K., Meguro, T., 2010b. Modeling user satisfaction transitions in dialogues from overall ratings, in: *Proceedings of the SIGDIAL 2010 Conference*, Association for Computational Linguistics, Tokyo, Japan. pp. 18–27.
- Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. *Information Theory, IEEE Transactions on* 14, 55 – 63.
- International Telecommunication Union, 2005. *Parameters Describing the Interaction with Spoken Dialogue Systems*. ITU-T Recommendation Supplement 24 to P-Series. International Telecommunication Union. Geneva, Switzerland. Based on ITU-T Contr. COM 12-17 (2009).
- ISO, 1998. *Ergonomic requirements for office work with visual display terminals (VDTs)*, Part 11: Guidance on usability. International Standardization Organization (ISO).
- Ives, B., Olson, M.H., Baroudi, J.J., 1983. The measurement of user information satisfaction. *Commun. ACM* 26, 785–793.
- Kamm, C.A., Litman, D., Walker, M., 1998. From novice to expert: The effect of tutorials on user expertise with spoken dialogue systems, in: *In Proceedings of the International Conference on Spoken Language Processing, ICSLP98*, pp. 1211–1214.
- Kim, W., 2007. Online call quality monitoring for automating agent-based call centers, in: *Proc. of the International Conference on Speech and Language Processing (ICSLP)*.
- Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.
- Langkilde, I., Walker, M., Wright, J., Gorin, A., Litman, D., 1999. Automatic prediction of problematic human-computer dialogues in how may i help you, in: *In Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU99*, pp. 369–372.
- Larcker, D., Lessig, V.P., 1980. Perceived usefulness of information: A psychometric examination. *Decision Sciences* , 121134.
- Levin, E., Pieraccini, R., 2006. Value-based optimal decision for dialog systems, in: *Proc. of Spoken Language Technology Workshop 2006*, pp. 198–201.
- Lindgaard, G., Dudek, C., 2003. What is this evasive beast we call user satisfaction? *Interacting with Computers* 15, 429–452.
- Litman, D., Hirschberg, J.B., Swerts, M., 2000. Predicting automatic speech recognition performance using prosodic cues, in: *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. pp. 218–225.
- Litman, D., Pan, S., 1999. Empirically evaluating an adaptable spoken dialogue system, in: *In Proceedings of the 7th International Conference on User Modeling*, pp. 55–64.
- Litman, D., Pan, S., 2002. Designing and evaluating an adaptive spoken dialogue system. *User Modeling and User-Adapted Interaction* 12, 111–137.
- Litman, D., Walker, M., Kearns, M.S., 1999. Automatic detection of poor speech recognition at the dialogue level, in: *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, Association for Computational Linguistics, Morristown, NJ, USA. pp. 309–316.
- Möller, S., 2005a. Neue itu-t-empfehlungen zur evaluierung telefonbasierter sprachdialogdienste, in: *Cremers, A.B., Manthey, R., Martini, P., Steinhage, V. (Eds.), INFORMATIK 2005 - Informatik LIVE! Band 2, Beiträge der 35. Jahrestagung der Gesellschaft für Informatik e.V. (GI)*, Bonn, Gesellschaft für Informatik.
- Möller, S., 2005b. *Quality of Telephone-based Spoken Dialogue Systems*. Springer, New York.
- Möller, S., Engelbrecht, K.P., Kühnel, C., Wechsung, I., Weiss, B., 2009. A taxonomy of quality of service and quality of experience of multimodal human-machine interaction, in: *Quality of Multimedia Experience, 2009. QoMEX 2009. International Workshop on*, pp. 7–12.
- Möller, S., Engelbrecht, K.P., Schleicher, R., 2008. Predicting the quality and usability of spoken dialogue services. *Speech Commun.* 50, 730–744.
- Paek, T., Horvitz, E., 2004. Optimizing automated call routing by integrating spoken dialog models with queuing models, in: *HLT-NAACL*, pp. 41–48.
- Platt, J.C., 1999. *Fast training of support vector machines using sequential minimal optimization*. MIT Press, Cambridge, MA, USA. pp. 185–208.
- Polzehl, T., Schmitt, A., Metze, F., 2010. Approaching multilingual emotion recognition from speech - on language dependency of acoustic/prosodic features for anger detection, in: *Proc. of the*

- Fifth International Conference on Speech Prosody, 2010. Speech Prosody 2010.
- Raux, A., Eskenazi, M., 2004. Non-native users in the let's go!! spoken dialogue system: Dealing with linguistic mismatch, in: HLT-NAACL, pp. 217–224.
- Riccardi, G., Gorin, A.L., Ieee, M., 2000. Stochastic language adaptation over time and state in natural spoken dialogue systems.
- Schmitt, A., Heinroth, T., Liscombe, J., 2009. On nomatches, noinputs and bargeins: Do non-acoustic features support anger detection?, in: Proceedings of the 10th Annual SIGDIAL Meeting on Discourse and Dialogue, SigDial Conference 2009, Association for Computational Linguistics, London (UK).
- Schmitt, A., Pieraccini, R., Polzehl, T., 2010a. Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics. Springer. chapter 'For Heavens sake, gimme a live person!' Designing Emotion-Detection Customer Care Voice Applications in Automated Call Cent.
- Schmitt, A., Polzehl, T., Minker, W., 2010b. Facing reality: Simulating deployment of anger recognition in ivr systems, in: Second International Workshop on Spoken Dialogue Systems (IWSDS).
- Schmitt, A., Schatz, B., Minker, W., 2011. Modeling and predicting quality in spoken human-computer interaction, in: Proceedings of the SIGDIAL 2011 Conference, Association for Computational Linguistics, Portland, Oregon, USA.
- Spearman, C., 1904. The proof and measurement of association between two things. *American Journal of Psychology* 15, 88–103.
- Walker, M., Kamm, C., Litman, D., 2000. Towards developing general models of usability with paradise. *Nat. Lang. Eng.* 6, 363–377.
- Walker, M., Langkilde-Geary, I., Hastie, H.W., Wright, J., Gorin, A., 2002. Automatically training a problematic dialogue predictor for a spoken dialogue system. *Journal of Artificial Intelligence Research*, 293–319.
- Walker, M., Litman, D., Kamm, A.A., Abella, A., 1998. Evaluating spoken dialogue agents with PARADISE: Two case studies. *Computer Speech and Language* 12.
- Walker, M., Litman, D., Kamm, C.A., Abella, A., 1997. Paradise: a framework for evaluating spoken dialogue agents, in: Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, Association for Computational Linguistics, Morristown, NJ, USA. pp. 271–280.
- Witten, I.H., Frank, E., 2005. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, San Francisco. 2nd edition edition.
- Zweig, G., Siohan, O., Saon, G., Ramabhadran, B., Povey, D., Mangu, L., Kingsbury, B., 2006. Automated quality monitoring in the call center with asr and maximum entropy, in: Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2006, pp. I–I.