NATURAL HUMAN ROBOT COMMUNICATION

Christian Fügen, Petra Gieselmann, Hartwig Holzapfel, Florian Kraft

Universität Karlsruhe (TH)
Institut für Theoretische Informatik, Interactive Systems Labs
Adenauerring 4, 76131 Karlsruhe, Germany
{fuegen|petra|hartwig|fkraft}@ira.uka.de

ABSTRACT

In this paper, we present work done over the past few years in our lab on speech, language and dialogue processing for the German collaborative research centre SFB-588 about "Humanoid Robots – Learning and Cooperating Multimodal Robots" to improve human-robot communication. Dialogues between humans and robots should be as natural as possible, wherefore we've analysed in which situations we have problems to understand the user in human-robot dialogues and built a reinforcement learning approach with a simulated user model, whereby we could improve naturalness and robustness significantly compared to a hand-crafted dialogue strategy. We also conducted experiments on proactive initiation of dialogues by the robot, the recognition of previously unknown names, and faces of persons. Another important issue is the interface between dialogue management and speech recognition. Here we could improve both, recognition and understanding performance by a tighter coupling of both components. While we participated in many evaluations on far distant speech recognition, we also did first experiments in sound event classification, to distinguish speech from other acoustic events, which become extremely important when moving from close to far distant speech recognition.

1. INTRODUCTION

Systems like telephony based information and booking systems with simple human-machine interfaces consisting of automatic speech recognition (ASR), dialogue management (DM) and speech synthesis are well known in the public and widely used. Nevertheless, those systems are restricted to limited domains and mostly support only a system initiated dialogue strategy in which the user is allowed to answer only to questions in a very limited way. Experiences with such systems show that while the performance is often acceptable, user acceptance is still a big problem. In the upcoming field of humanoid and human-friendly robots user acceptance is even more an issue. Therefore, the ability of the robot to interact in a simple, unconstrained and natural way with its users is of great importance. A user should neither be restricted by a command based speech interface nor should he be forced to wear head-mounted microphones in order to communicate with the robot. Instead, spontaneous, mixed initiative speech dialogues recorded by distant microphones should be possible.

In the German collaborative research centre SFB-588 about "Humanoid Robots – Learning and Cooperating Multimodal Robots" [1] our lab is working on improving the naturalness of human-robot communication. Therefore we present in this paper work done over the past few years in our lab on speech, language and dialogue processing in the field of human-robot communication. We analysed in which situations conventional dialogue systems have problems to understand the user in human-robot dialogues and built a reinforcement learning approach with a simulated user model, whereby we could improve naturalness and robustness significantly compared to a hand-crafted dialogue strategy. We also conducted experiments on proactive initiation of dialogues by the robot. Another important research area is the recognition of previously unknown words. Both, the speech recognition and the dialogue system should be able to detect and to react properly in such situations. First experiments were done in this direction, by recognising previously unknown names and faces of persons. To improve both, the recognition and understanding performance, we implemented a tighter coupling between the speech recognition and the dialogue management. This allows us to direct the decoding process depending on the dialogue state, so that areas of the speech recogniser's search space are favoured while others are penalised. While we participated in many evaluations on far distant speech recognition, we also did first experiments in sound event classification, to distinguish speech from other acoustic events, which become extremely important when moving from close to far distant speech recognition.

2. DIALOGUE MANAGEMENT

The dialogue manager interprets multimodal input and generates responses and spoken output. Its strategy defines how communication with the user is directed and interpreted. Within the collaborative research centre SFB-588, we see the main challenge for dialogue management in providing a natural way to communicate and interact with the robot, and provide the robot with an interface to the environment through which it can obtain new information and learn from communication partners. Natural communication with the robot includes multimodal interaction as well as robust error-tolerant dialogue strategies. For the latter we present two approaches in this paper, one by addressing errors in human-robot dialogues, and the second approach by providing the robot to learn and adapt its dialogue strategies. We furthermore present work on proactive initiation of conversations, work on acquiring information about persons, and learning new words of objects and their meanings.

2.1. Error Analysis of Human-Robot Dialogues

One of the biggest challenges in current human-robot communication are misunderstandings and problematic situations: Errors and misunderstandings often result in error spirals from which the user can hardly escape which leads to user frustration and task failure. These errors are on one hand due to imperfect speech understanding but on the other hand also to user uncertainty about how to interact with a robot. Within the present study, we focus on the second point and work towards a new generation strategy which includes detecting problematic situations and helping the user.

We made a preliminary user study to get a deeper insight into the dialogue problems and classified them from the user's point of view:

- The user is unsure what to do.
- The robot's response is inconsistent with the user's expectation.
- There is no response from the robot to a user utterance.
- The robot tells the same information for several times.

All these problems result in user frustration and task failure. In the first case, the user explicitly asks for help. Therefore, we invented a general help strategy which covers all the tasks the robot can accomplish in a hierarchical structure so that the user first gets general and then more and more fine-grained information about the robot's capabilities. The robot also explicitly hints the user to tasks it cannot do to avoid errors caused by unknown vocabulary words. For example, all the users in the present study asked the robot to clean, but since it was unable, this resulted in recognition errors and misunderstandings.

In the other cases, the user feels lost and somehow confused by the dialogue. In the user study, we find different factors indicating that the user needs help and that a problem occurred:

- No speech act can be found in the user utterance.
- The user utterance is inconsistent with the current discourse.
- The user utterance can only be partly parsed.
- The user utterance is inconsistent with the robot's expectations.
- The user asks for the same information for several times.

In addition, certain wordings indicate that the user tries to correct the preceding utterance, such as "no", "I do not want". All these features are used for an automatic problem detection: We define an error correction necessity which increases when one of the above features can be found in a user utterance. When the error correction necessity is above a given threshold, the robot explains its capabilities and the discourse information is finally cleared so that the user can start from scratch again. When the user is then able to continue his dialogue with the robot without problems, the error correction necessity decreases from turn to turn, but stays at a certain level indicating that there were problems within this dialogue. In this way, we can avoid error spirals and support the user in problematic situations. Our analysis of the user study showed that a large number of user utterances can be classified reliably using these features. In the future, we might also include prosodic-acoustic features because they have been proven useful for automatic error detection as shown by [2].

We keep track of the user knowledge in a user model so that the user does not get the same information twice, if not explicitly asking for it. Whenever a user talks for the second time with the robot, the robot knows what it already told him and therefore we can classify a user utterance asking for already known information also as problematic. In the future, we will evaluate the problem detection module in combination with the user model and define useful thresholds for starting problem correction dialogues.

2.2. Learning in Human-Robot Dialogues

Human-Robot dialogues are by nature multimodal and the robot's perception includes a variety of sensor signals that deliver important information to the dialogue manager. We use such non-speech information to obtain information about the environment and about the user. Using a person-tracking module the dialogue manager gets notified about pointing gestures from the user and can resolve this information in the context of spoken utterances to communicate with the user in a very natural way. Multimodal fusion as well as dialogue management is performed by the dialogue manager TAPAS [3] that has explicitly been developed for this purpose. It uses the same goal-based dialogue framework as the ARIADNE [4] system. For fusion of speech and pointing gestures we have developed a robust rule-based approach which is especially suitable to handle input with many false detections. Fusion of speech and gesture events observes their semantic content as well as time constraints and allows processing of n-best lists. In experiments, the system showed a significant improvement by being able to sort out falsely detected gesture events [5].

In our approach for multimodal fusion, the robot needs to know the exact position of objects that are usually not in the visual field of the robot, to calculate possible matches with the pointing gesture of the person. It is in this case also restricted to known objects, with a given wording and their semantic representation. To overcome this limitation, we have started to explore online learning mechanisms that allow the robot to explore and learn about its environment during an interaction. First, the system needs to detect its limitations that are given for example when the person references objects that are not known to the robot. We have conducted experiments to test the suitability of detecting unknown words in speech, to clarify references to unknown objects and then to learn their semantic meaning. Within an evaluation set of 42 dialogues, with eleven test persons, where persons requested unknown objects, 10 dialogues were aborted after recognising the wrong speech. Of the remaining 32 dialogues the unknown word was detected in 84% on the first utterance and in the remaining 16% after a clarification question. 75% of these 32 dialogues lead to successfully learning the new word, by adding it to the correct position in the ontology and to the grammar. Being able to adapt to a changing environment and being able to adapt to unforeseen situations will play a major role in the acceptability and success of an autonomous humanoid robot.

In more recent experiments we have used the same mechanism to detect unknown words with dialogue strategies to learn names of persons and combine them with a visual ID [6]. There, the robot is the initiator of interactions, and tries to obtain the attention from persons by trying to engage them into a conversation. Different actions have been tested, like playing sounds, turning the robot-head to the person, and inviting the person to talk to the system. The experiments have shown that different actions can be used to obtain the attention from a person, however, only speech was suitable to initiate a conversation. By a combining visual contact and speech, the robot was finally able to obtain attention and initiate a conversation with persons. Following the initiation of a dialogue, the robot can then ask persons about their names and store this information to recognise the person in future encounters. Table 1 shows the evaluation of different actions with subjective measures how much they influence attention, and how much they are suitable to initiate a conversation. The experiment was conducted with eleven persons that were instructed to walk by the robot and judge its actions. Each user had to do this five times, each iteration with different combination of actions. The evaluated categories are 'eye-catching' (does the action influence my attention?) and 'suitable' (do you feel you should start a dialogue with the system?). The values for eye-catching are scaled to 0 (no influence), 1 (medium), 2 (annoying). The values for 'suitable' are scaled to -1 (not suitable), 0 (a little bit), 1 (yes).

The final evaluation then aims to evaluate how well the robot was able to initiate a dialogue. It was conducted in 100 attempts distributed over five persons. The experiment was artificial in a way that the persons re-did the same experiment a couple of times and decided for themselves if they would interact with the robot. The absolute numbers are thus subjective to the willingness of the persons to interact with the robot. During some of the iterations the users could not be tracked correctly, e.g. due to changing light conditions. When the system failed to track the user no interaction could be initiated. Table 2 shows, for each user, the tracker-recognition rate and the success rate to start a dialogue. The table shows a dependency of recognition rate to success rate, but also a bias in a different behaviour by different users.

Table 1. Evaluation of different system actions.

action	eye-catching	suitable
play sound	0.9	-0.3
turn head	0.9	-0.3
say 'hello'	0.9	0.8
play sound then say 'hello'	0.9	0.3
turn head then say 'hello'	1.0	0.9

Table 2. Evaluation of the success rates per user.

person no.	recognition rate	success rate
1	80%	35%
2	100%	85%
3	60%	30%
4	70%	15%
5	100%	50%

In a related setup we have started to explore learning methods, especially reinforcement learning, that allow the robot to learn and adapt its dialogue strategy. So far, different reinforcement learning approaches have been studied to learn a dialogue strategy either from an existing dialogue corpus, from online experience, or within a simulation environment. Our setup describes the strategy of a bartender robot that learns in a multimodal scenario, to request information which object to serve [7]. The user has the option to describe the desired object by speech or by pointing at the object. In addition to interpreting unconstrained input by the user, the dialogue engine can decide among the options to request information about the type, the location or the colour of an object, confirm each single information slot, the complete information collected so far, or point at a single object to confirm the object itself. Since reinforcement learning requires a large amount of data, we chose to create a user simulation, so that training the strategy could be accomplished completely within the simulation. The simulation included a user model with speech utterances and pointing gestures, as well as a stochastic error model for each input modality. The chosen approach was the first to combine multimodal input with error models for simulation-based reinforcement learning. The results show that training the simulation from a small pre-collected Wizard-of-OZ study, already led to a very accurate dialogue model. The results of a final evaluation with the trained dialogue strategy showed a significant improvement

Table 3. Word error rates (WER) for single and multiple distant microphones, e.g. table tops and microphone arrays on lectures. Evaluation data for 2006 was significantly more difficult than compared to 2005

	2004	2005	2006
single	75.1%	66.5%	54.7%
multiple	69.6%	55.8%	53.4%

over a non-trivial hand-crafted dialogue strategy.

3. ROBUST SPEECH RECOGNITION

Moving away from head-mounted microphones to far distant microphones is another important issue when designing a natural human-machine interface. The robot should be able to react all the time to events occurring in his environment, which means, that his sensors – microphones as ears and cameras as eyes – have to be turned on all the time.

In this Section, we focus on the ears of the robot and will present techniques and first experimental results bringing us closer towards a natural human-machine interface, namely sound event classification [8] and a tighter coupling between the speech recogniser and the dialogue manager [9, 10].

To reduce the performance gap between far distant and close talking speech recognition we participated also in evaluations organised by the CHIL project [11] or the National Institute of Standards and Technology (NIST). The evaluations dealt with distant speech generally, wherefore meetings and lectures, recorded with different kinds of microphones have to be recognised. Since June 2004 we could improve the speech recognition word error rate (WER) on Lectures on far distant speech from 69.6% to 53.4% (see Table 3) [12]. Nevertheless, results obtained by the evaluations can be used to improve the speech recognition also in limited domains.

All speech recognition experiments were done with the help of the Janus Recognition Toolkit (JRTk) featuring the IBIS decoder [13], developed at our lab.

3.1. Sound Event Classification

Due to the fact, that the robot's microphones are always open, we need to detect and classify all environmental sounds and have to connect them with actions by the robot. Speech signals for example should be transfered to the speech recogniser, while a door bell ring or a microwave beep should be handled elsewhere. This is not an easy problem, because speech and noises from the environment and from the robot itself can interfere with each other.

In this Section we focus on sound event classification in a kitchen environment, because in the main SFB-588 scenario the humanoid robot is intended to assist elderly or disabled humans in kitchen tasks such as cooking, cleaning and to provide safety assurance. The ability to detect important kitchen sounds is vital to this set of functions and can improve the recognition of far distant speech in such noisy environments; many important state indicators in the kitchen, like alarms, bells, buzzers, water boiling, or oil beginning to sizzle in a pan, leave little or no visual evidence. Towards the goal of distinguishing these sound events, we developed a novel feature extraction method [8]. Our method learns Independent Component Analysis (ICA) basis functions [14] over multi-frame windows of frequency domain features to capture inter-frame temporal dependencies.

For the kitchen domain we found relevant sound classes according to the following three categories:

- observation of dangerous situations
- observation of human activities
- observation of automated activity

Using a Sony ECM-719 microphone roughly 6000 instances of real-world kitchen sounds had been collected which are distributed over 21 classes. These instances were divided at random into a training (70%) and a test set (30%) as shown in Table 4.

3.1.1. Experimental Results

For the baseline system BASE we used MFCCs with 20ms windows and 10ms shifts. When adding first and second temporal derivatives the feature vectors resulted in a 39-dimensional feature space. For the test systems we derived 20 log mel spectra from the power spectra. The dimensionality was reduced to 13 dimensions for the test system's features using a PCA transformation which allowed us to retain at least 95% of the total eigenvalue mass. Finally a global ICA transformation [15] was applied which had been trained on single (ICA1) and multi-frame (ICA7) feature vectors (3 frames left and 3 frames right context) for all classes.

After training GMMs and ergodic 3-state HMMs, we evaluated the models using the maximum likelihood criterion. For all systems the average per-class error and the average per-class precision were computed.

For a fair comparison the number of Gaussians was kept fixed at 15 (i.e. for the 3-state HMMs the Gaussians were distributed evenly among the states). Note that this actually means that there are three times as many parameters in the baseline systems as for the ICA systems. Table 5 shows that for both model types the ICA systems outperform the baseline significantly. Further a gain can be observed for ICA systems when using temporal ICA basis functions. These temporal basis functions are able to cover both frequency information and temporal context at a time as shown in figure 1. Note that there is a performance difference between GMMs and HMMs which is basically due to the constraint of having exact 5 Gaussians per state for the HMMs. When relaxing this constraint to fixing only the total number of Gaussians to 15 per model, the corresponding ergodic HMM based ICA7 system gives an only slightly worse error of 9.4% compared to

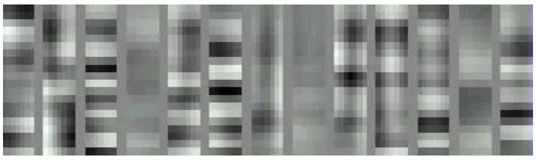

Fig. 1. Seven-Frame ICA Basis Functions

Table 4. Sample counts and durations in seconds per class.

class	training (duration)	test (duration)	total (duration)
boiling	221 (662)	98 (319)	319 (981)
bread cutter	25 (40)	11 (27)	36 (67)
cutting vegetables	134 (89)	58 (41)	192 (130)
door	114 (101)	50 (44)	164 (144)
door bell	50 (110)	22 (55)	72 (164)
egg timer ring	11 (34)	6 (17)	17 (51)
footsteps	240 (140)	104 (66)	344 (206)
lighter	84 (42)	37 (20)	121 (61)
match	141 (131)	62 (59)	203 (189)
microwave beep	110 (30)	49 (17)	159 (47)
others	858 (1130)	369 (547)	1277 (1677)
oven switch	472 (133)	208 (60)	680 (194)
oven timer	12 (16)	6 (8)	18 (24)
over-boiling	186 (129)	81 (70)	267 (199)
pan stove	584 (308)	256 (132)	840 (439)
pan sizzling	107 (343)	46 (146)	153 (489)
telephone	134 (920)	63 (393)	197 (1313)
speech	125 (82)	55 (38)	180 (120)
stove error	18 (12)	8 (5)	26 (17)
toaster	119 (92)	53 (46)	172 (138)
water	421 (1129)	184 (464)	605 (1593)
total	4166 (5670)	1826 (2573)	5992 (8243)

9.2%. We further suspect to get better results with HMMs when choosing class dependent topologies.

3.2. Dialogue-Context Dependent Speech Recognition

Our goals for a tight coupling between a speech recogniser and a dialogue manager are to share as much information between these two components as possible, to improve especially far distant speech recognition and hence system understanding performance. Therefore, the implementations of IBIS [13] and TAPAS allow us to share the linguistic knowledge sources, i.e. context-free grammars, which gives us the ability to use the results of one component directly for improving the performance of the other component in the next step [9].

Table 5. Error and precision, number of Gaussians fixed at 15 per class

System	GMM		ERG3	
	Error	Precision	Error	Precision
BASE	12.4%	80.6%	12.2%	82.8%
ICA1	10.6%	82.8%	10.9%	82.2%
ICA7	**9.2%**	**85.0%**	**10.2%**	**83.4%**

Due to the fact, that IBIS uses linked recursive transition networks (RTNs) for its internal grammar representation, the original grammar structure can be directly accessed, which has several advantages:

- IBIS can also be used as a parser for natural language processing. Therefore, a separate parser is superfluous. The recognised and parsed output can be directly given to TAPAS.
- Rules can be activated/deactivated or weighted (penalised) during run-time, which can be used, e.g., to restrict the decoding process to sub-grammar parts only.

We have divided all entry rules of the grammars into two sets, a ResponseSet consisting only of rules likely to be used in between a dialogue, i.e. elliptical expressions and responses to information requests and a QuerySet consisting of rules likely to be used at the beginning of a dialogue. TAPAS uses a generic approach for an expectation model that describes which utterances are most likely used by the user for his next query/response and gives this information to IBIS [10]. The expectation model is created based on the current dialogue context and expected information. At the beginning of the dialogue, all rules of the ResponseSet are penalised, whereas during the dialogue the specified set of rules out of the ResponseSet given by TAPAS are preferred over all others. It should be emphasised that still other user inputs can be recognised by IBIS which is conform to a mixed initiative dialogue system.

3.2.1. Experimental Results

We compared the speech recognition results, i.e. the word error rates (WER) and sentence error rates (SER) in the SFB domain of a household robot on both, close and distant talking microphones, to measure the difference in performance gain by introducing the new methods. Therefore, we collected a dialogues of 8 different speakers which consists of 346 spontaneous speech queries and 300 responses to clarification questions from the robot. For the distant data only one microphone at a distance of about 2-3m from the speaker was used, which means that no array processing could be done to improve the speech recognition results. Given the parsed transcripts of the pre-recorded dialogs, the dialog manager was used to compute the preferred rules for the next user response depending on the dialog context. The weighting parameters for the grammar sets were optimised on a cross validation set.

In Table 6 the baseline results are reported. It can be seen that the recognition results for the user responses are worse than for the user queries. Especially for the distant condition the WERs for the user responses are about 50% worse than for the user queries. The sentence error rates do not vary as much as the word error rates. When using our context dependent grammar weighting as described above it can be seen

Table 6. Close (C) and distance (D) talking word and sentence error rates (baselines).

	WER	SER
User queries (C)	20.21%	34.10%
User responses (C)	30.28%	30.67%
Overall (C)	23.52%	32.51%
User queries (D)	30.53%	51.15%
User responses (D)	43.77%	43.62%
Overall (D)	34.86%	47.66%

Table 7. Close (C) and distant (D) talking word and sentence error rates together with their relative improvements compared to Table 6.

			improvement	
	WER	SER	WER	SER
Queries (C)	19.63%	33.53%	2.87%	1.67%
Responses (C)	29.11%	30.00%	3.86%	2.18%
Overall (C)	22.74%	31.89%	3.32%	1.91%
Queries (D)	28.81%	50.29%	5.63%	1.68%
Responses (D)	36.77%	39.60%	15.99%	9.22%
Overall (D)	31.41%	45.33%	9.90%	4.89%

in Table 7 that there is an overall reduction of the WER of 3.3% for the close and 9.9% for the distant talking condition. Whereas there is a smaller gain for the user queries, the user responses are recognised much better. It can also be seen, that the relative improvement increases for the distant condition.

4. CONCLUSION

To improve the human-robot communication, we analysed problematic communication situations and evaluated possible reasons for errors within the dialogues. We observed the user behaviour whenever the robot tries to initiate a dialogue and developed methods how the robot can start a new conversation. In addition, we extended the dialogue module so that the robot can learn and adapt its strategy to the current situation. We developed methods to detect and recognise new words so that the robot can also adapt to new environments. For a natural interaction with the robot it is important that the user can talk to it without head-mounted microphones. Therefore, we developed a sound-event classifier which can distinguish different kitchen sounds on one hand and speech on the other hand. To further improve the recognition accuracy and the understanding rate, we used context-dependent speech recognition.

5. ACKNOWLEDGEMENTS

This research is partially funded by the German Research Foundation (DFG) under Sonderforschungsbereich 588: "Humanoid Robots – Learning and Cooperating Multimodal Robots".

6. REFERENCES

[1] "Sonderforschungsbereich 588: Humanoide Roboter - Lernende und kooperierende multimodale Roboter," http://www.sfb588.uni-karlsruhe.de.

[2] G. A. Levow, "Characterizing and recognizing spoken corrections in human-computer dialogue," *Proc. of the ACL*, 1998.

[3] H. Holzapfel, "Towards Development of Multilingual Spoken Dialogue Systems," in *Proc. of the 2nd Language and Technology Conference (L&T'05)*, Poznan, 2005.

[4] M. Denecke, "Rapid Prototyping for Spoken Dialogue Systems," in *Proc. of the 19th International Conference on Computational Linguistics*, Taiwan, 2002.

[5] H. Holzapfel, K. Nickel, and R. Stiefelhagen, "Implementation and Evaluation of a Constraint-Based Multimodal Fusion System for Speech and 3D Pointing Gestures," in *Proc. of the International Conference on Multimodal Interfaces*, State College, PA, 2004.

[6] H. Holzapfel, T. Schaaf, H. K. Ekenel, C. Schaa, and A. Waibel, "A Robot learns to know people - First Contacts of a Robot," *Springer Lecture Notes in Artificial Intelligence, KI 2006, to appear*, 2006.

[7] T. Prommer, H. Holzapfel, and A. Waibel, "Rapid Simulation-Driven Reinforcement Learning of Multimodal Dialog Strategies in Human-Robot Interaction," in *INTERSPEECH*, Pittsburgh, PA, USA, 2006.

[8] F. Kraft, R. Malkin, T. Schaaf, and A. Waibel, "Temporal ICA for Classification of Acoustic Events in a Kitchen Environment," in *INTERSPEECH*, Lisbon, Portugal, 2005.

[9] C. Fügen, H. Holzapfel, and A. Waibel, "Tight Coupling of Speech Recognition and Dialog Management – Dialog-Context Dependent Grammar Weighting for Speech Recognition," in *ICSLP*, Jeju-Islands, Korea, 2004.

[10] H. Holzapfel and A. Waibel, "A Multilingual Expectations Model for Contextual Utterances in Mixed-Initiative Spoken Dialogue," in *INTERSPEECH*, Pittsburgh, PA, USA, 2006.

[11] "Computers in the Human Interaction Loop," http://chil.server.de.

[12] C. Fügen, M. Wölfel, J. W. McDonough, S. Ikbal, F. Kraft, K. Laskowski, M. Ostendorf, S. Stüker, and K. Kumatani, "Advances in Lecture Recognition: The ISL RT-06S Evaluation System," in *INTERSPEECH*, Pittsburgh, PA, USA, 2006.

[13] H. Soltau, F. Metze, C. Fügen, and A. Waibel, "A One Pass-Decoder Based on Polymorphic Linguistic Context Assignment," in *ASRU*, Trento, Italy, 2001.

[14] A. J. Bell and T. J. Sejnowksi, "An informationmaximisation approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, 1995.

[15] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, 1997.