7. ACKNOWLEDGEMENTS

The authors wish to thank all members of the Interactive Systems Laboratories, especially P. Geutner, T. Kemp, T. Sloboda, M. Woszczyna, Laura L. Mayfield, and Puming Zhan for useful discussions and active support.

8. REFERENCES

[1] Y.K. Muthusamy, R.A. Cole, and B.T. Oshika: *The OGI multi-language telephone speech corpus*. Proceedings of the ICSLP 1992.

[2] Y.K. Muthusamy, E. Barnard, and R.A. Cole: *Reviewing Automatic Language Identification*. IEEE Signal Processing Magazine, Vol. 11 No. 4 Oktober 1994, pp. 33-41.

[3] M. Woszczyna, N. Aoki-Waibel, F.D. Buø, N. Coccaro, K. Horiguchi, T. Kemp, A. Lavie, A. McNair, T. Polzin, I. Rogina, C.P. Rose, T. Schultz, B. Suhm, M. Tomita, A. Waibel: *JANUS 93: Towards Spontaneous Speech Translation*. Proceedings of the ICASSP 1994, volume 1, pp. 345-348.

[4] M.A. Zissmann and E. Singer: *Automatic Language Identification of Telephone Speech Messages using Phoneme Recognition and N-gram Modeling*. Proceedings of the ICASSP 1993, volume 2, pp. 309-402.

[5] L.F. Lamel and J. Gauvain: *Identifying Non-linguistic Speech Features*. Proceedings of the Eurospeech 1993, volume 1, pp. 23-30.

[6] Y. Muthusamy, K. Berkling, T. Arai, R.A. Cole, and E. Barnard: *Comparison of Approaches to Automatic Language Identification using Telephone Speech*. Proceedings of ...eech 1993, pp. 1307-1310.

... Reyes, T. Seino, and S. Nakagawa ... *e Language Identification Methods based* ... of the ICSLP 1994,

... ultz, and I. Rogina: *Acoustic and* ... *ing of Human and Nonhu-*

... and V.W. Zue: *Automatic Language Identification using a Segment-based Approach*. Proceedings of the Eurospeech 1993, pp. 1303-1306.

[9] B. Suhm, P. Geutner, T. Kemp, A. Lavie, L. Mayfield, A.E. McNair, I. Rogina, T. Schultz, T. Sloboda, W. Ward, M. Woszczyna, A. Waibel: *JANUS: Towards Multilingual Spoken Language Translation*. DARPA Speech and Natural Language Workshop 1994.

[10] K.M. Berkling, T. Arai, and E. Barnard: *Analysis of Phoneme-based Features for Language Identification*. Proceedings of the ..., volume 1, pp. 289-292.

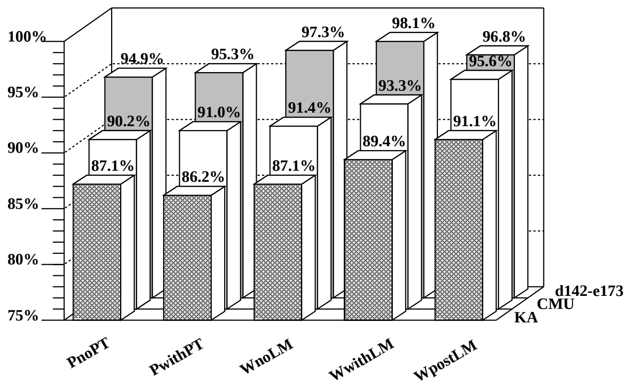

Figure 5: Comparison of the five systems

incorporation of knowledge sources improves the language identification accuracy significantly. For the cross-channel tests system WpostLM leads to best results. In all cases the performance increases when adding the dictionary. Furthermore, tests including the language-dependent word grammars outperform the results of those without linguisti
knowledge. Testing under different channel
tions increases the performance si
the two-stage system Wp o
impact. Note how
guage iden
are

5. LID USING DIFFERENT KNOWLEDGE SOURCES

For each language we constructed five systems applying different levels of knowledge.

5.1. System PnoPT

PnoPT is a recognizer with phoneme-based acoustic modeling. For each language a system with context-independent phonemes which are modeled by SCHMMs with 50 tied mixture weights was build. For the German language we used a set of 46 phonemes, for English 54 phonemes and for Spanish 48 phonemes. The phoneme sets include special noise models to model human and nonhuman noises as described in [11].

5.2. System PwithPT

PwithPT is similar to P
tactics i.e.
phonol

the increasing ambiguity when adding languages to be identified to the system. We use a parallel architecture as shown in figure 2.

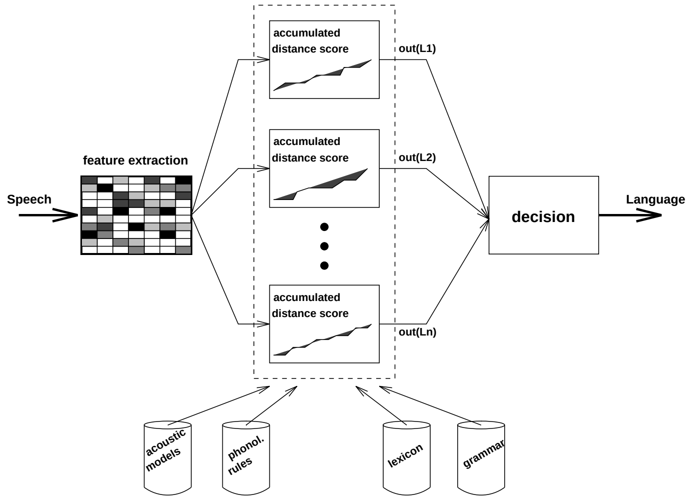

Figure 2: parallel architecture of our LID system

4. CROSS-CHANNEL CONTROL

As mentioned before, many constraints are applied to the data collection procedure to guarantee the comparability of the recorded data. Both of the sites at which we collect our data use the same closed speaker microphones, the same hardware
to digitize the speech input, and the sa
nario and calendar. To control for
nel variations or different
have recorded addi
collect Germ
the U

and Karlsruhe University over the last 20 months.
In each session, two people are asked to schedule
a meeting with their dialog partners. Constraints
for the scenario, the calendar and the collection
procedures of the data guarantee the comparabil-
ity of the data recorded at different sites. The col-
lection scenario and requirements are described in
detail in [3] and [9]. The SST corpus current
sists of dialogs in the languages En
Spanish, and Korean spontan
tive speakers. The col
has also begu
current

EXPERIMENTS WITH LVCSR BASED LANGUAGE IDENTIFICATION

T.Schultz, I.Rogina, and A. Waibel

Interactive Systems Laboratories
University of Karlsruhe (Germany), Carnegie Mellon University (USA)
{tanja,rogina,waibel}@ira.uka.de

ABSTRACT

Automatic language identification is an important problem in building multilingual speech recognition and understanding systems. We have developed a front-end LID module based on LVCSR to identify English, German, and Spanish language for use in spontaneous speech-to-speech translation. We studied the constitution of different levels of knowledge to identify a language, i.e. the phonetic, phonotactic, lexical, and syntactic-semantic knowledge. A comparison of LID systems usi
different levels of these knowledge sour
sented. We showed that the inc
ical and linguistic know
of the language

In recent
re