

Acknowledgments. The authors would like to thank Sven Magg, Erik Strahl and the students of the international Master Intelligent Adaptive Systems for their help in collecting the SPONT corpus. This research has been partly supported by the EU project RobotDoC under 235065 ROBOT-DOC from the 7th Framework Programme (FP7), by the DFG German Research Foundation (grant #1247) – International Research Training Group CINACS, and by the Daimler and Benz foundation (PostDoc grant #32-02/13).

References

- Baumann, T.; Atterer, M.; and Schlangen, D. 2009. Assessing and improving the performance of speech recognition for incremental systems. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 380–388.
- Bisani, M., and Ney, H. 2008. Joint-sequence models for grapheme-to-phoneme conversion. *Speech Communication* 50(5):434–451.
- Garofolo, J. S.; Lamel, L. F.; Fisher, W. M.; Fiscus, J. G.; Pallett, D. S.; and Dahlgren, N. L. 1993. DARPA TIMIT acoustic phonetic continuous speech corpus CDROM.
- Garofolo, J.; Fiscus, J. G.; and Fisher, W. M. 1997. Design and preparation of the 1996 HUB-4 broadcast news benchmark test corpora. In *Proceedings of the 1997 DARPA Speech Recognition Workshop*, 15–21.
- Heinrich, S., and Wermter, S. 2011. Towards robust speech recognition for human-robot interaction. In Narioka, K.; Nagai, Y.; Asada, M.; and Ishiguro, H., eds., *Proceedings of the IROS2011 Workshop on Cognitive Neuroscience Robotics (CNR)*, 29–34.
- IPA, I. 1999. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.
- Jaitly, N.; Nguyen, P.; Senior, A. W.; and Vanhoucke, V. 2012. Application of pretrained deep neural networks to large vocabulary speech recognition. In *Proceedings of the 13th Annual Conference of the International Speech Communication Association (Interspeech)*. ISCA.
- Jurafsky, D., and Martin, J. H. 2009. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Pearson International, 2 edition.
- Levenshtein, V. I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics – Doklady* 10(8):707–710.
- McGraw, I., and Gruenstein, A. 2012. Estimating word-stability during incremental speech recognition. In *Proceedings of the 13th Annual Conference of the International Speech Communication Association (Interspeech)*. ISCA.
- Mermelstein, P. 1976. Distance measures for speech recognition – psychological and instrumental. In *Proceedings of the Joint Workshop on Pattern Recognition and Artificial Intelligence*, 374–388.
- Millette, G., and Stroud, A. 2012. *Professional Android Sensor Programming*. ITPro collection. Wiley.
- Morbini, F.; Audhkhasi, K.; Artstein, R.; Van Segbroeck, M.; Sagae, K.; Georgiou, P.; Traum, D. R.; and Narayanan, S. 2012. A reranking approach for recognition and classification of speech input in conversational dialogue systems. In *Proceedings of the 2012 IEEE Spoken Language Technology Workshop (SLT)*, 49–54. IEEE.
- Morbini, F.; Audhkhasi, K.; Sagae, K.; Artstein, R.; Can, D.; Georgiou, P.; Narayanan, S.; Leuski, A.; and Traum, D. 2013. Which ASR should I choose for my dialogue system? In *Proceedings of the 14th annual SIGdial Meeting on Discourse and Dialogue*, 394–403.
- Rayner, M.; Carter, D.; Digalais, V.; and Price, P. 1994. Combining knowledge sources to reorder n-best speech hypothesis lists. In *Proceedings of the ARPA Human Language Technology Workshop*, 217–221.
- Schalkwyk, J.; Beeferman, D.; Beaufays, F.; Byrne, B.; Chelba, C.; Cohen, M.; Kamvar, M.; and Strophe, B. 2010. Your word is my command: Google search by voice: A case study. In *Advances in Speech Recognition*. Springer. 61–90.
- Twiefel, J. 2014. Development and evaluation of semantically constrained speech recognition architectures. Master’s thesis, Universität Hamburg, Dept. of Informatics, Hamburg, DE.
- Van Heerden, C.; Schalkwyk, J.; and Strophe, B. 2009. Language modeling for what-with-where on GOOG-411. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech)*. ISCA.
- Walker, W.; Lamere, P.; Kwok, P.; Raj, B.; Singh, R.; Gouvea, E.; Wolf, P.; and Woelfel, J. 2004. Sphinx-4: A Flexible Open Source Framework for Speech Recognition. Technical Report SMLI TR2004-0811, Sun Microsystems Inc.
- Young, S. J.; Russell, N.; and Thornton, J. 1989. Token passing: A simple conceptual model for connected speech recognition systems. Technical Report CUED/F-INFENG/TR, Cambridge University Engineering Department.
- Zgank, A., and Kacic, Z. 2012. Predicting the acoustic confusability between words for a speech recognition system using Levenshtein distance. *Electronics and Electrical Engineering* 18(8):81–84.
- Ziółko, B.; Gałka, J.; Jadczyk, T.; and Skurzok, D. 2010. Modified weighted Levenshtein distance in automatic speech recognition. In *Proceedings of the 16th Conference on Applications of Mathematics in Biology and Medicine*,