

References

- Abu-Mostafa, Y. S. (1989), "The Vapnik–Chervonenkis dimension: information versus complexity in learning," *Neural Computation* **1**, 312–317.
- Akaike, H. (1974), "A new look at the statistical model identification," *IEEE Transactions on Automatic Control* **19**(6), 716–723.
- Akaike, H. (1980), "Likelihood and the Bayes procedure," in J. M. Bernardo, M. H. DeGroot, D. V. Lindley & A. F. M. Smith, eds, *Bayesian Statistics*, University Press, Valencia, Spain, pp. 143–166.
- Akita, Y., & Kawahara, T. (2004), "Language model adaptation based on PLSA of topics and speakers," Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH), pp. 1045–1048.
- Aldous, D. (1985), "Exchangeability and related topics," *École d'Été de Probabilités de Saint-Flour XIII1983*, pp. 1–198.
- Anastasakos, T., McDonough, J., Schwartz, R., & Makhoul, J. (1996), "A compact model for speaker-adaptive training," Proceedings of International Conference on *Spoken Language Processing (ICSLP)*, pp. 1137–1140.
- Anguera Miro, X., Bozonnet, S., Evans, N., *et al.* (2012), "Speaker diarization: A review of recent research," *IEEE Transactions on Audio, Speech, and Language Processing* **20**(2), 356–370.
- Antoniak, C. E. (1974), "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems," *Annals of Statistics* **2**(6), 1152–1174.
- Attias, H. (1999), "Inferring parameters and structure of latent variable models by variational Bayes," Proceedings of the Conference on *Uncertainty in Artificial Intelligence (UAI)*, pp. 21–30.
- Axelrod, S., Gopinath, R., & Olsen, P. (2002), "Modeling with a subspace constraint on inverse covariance matrices," Proceedings of International Conference on *Spoken Language Processing (ICSLP)*, pp. 2177–2180.
- Bahl, L. R., Brown, P. F., de Souza, P. V., & Mercer, R. L. (1986), "Maximum mutual information estimation of hidden Markov model parameters for speech recognition," Proceedings of International Conference on *Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 49–52.
- Barber, D. (2012), *Bayesian Reasoning and Machine Learning*, Cambridge University Press.
- Barker, J., Vincent, E., Ma, N., Christensen, H., & Green, P. (2013), "The PASCAL CHiME speech separation and recognition challenge," *Computer Speech and Language* **27**, 621–633.
- Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970), "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *The Annals of Mathematical Statistics*, pp. 164–171.
- Beal, M. J. (2003), Variational algorithms for approximate Bayesian inference, PhD thesis, University of London.

- Beal, M. J., Ghahramani, Z., & Rasmussen, C. E. (2002), "The infinite hidden Markov model," *Advances in Neural Information Processing Systems* **14**, 577–584.
- Bellegarda, J. (2004), "Statistical language model adaptation: review and perspectives," *Speech Communication* **42**(1), 93–108.
- Bellegarda, J. R. (2000), "Exploiting latent semantic information in statistical language modeling," *Proceedings of the IEEE* **88**(8), 1279–1296.
- Bellegarda, J. R. (2002), "Fast update of latent semantic spaces using a linear transform framework," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 769–772.
- Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003), "A neural probabilistic language model," *Journal of Machine Learning Research* **3**, 1137–1155.
- Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis*, Second Edition, Springer-Verlag.
- Bernardo, J. M., & Smith, A. F. M. (2009), *Bayesian Theory*, Wiley.
- Berry, M. W., Dumais, S. T., & O'Brien, G. W. (1995), "Using linear algebra for intelligent information retrieval," *SIAM Review* **37**(4), 573–595.
- Bilmes, J. A. (1998), A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, Technical Report TR-97-021, International Computer Science Institute.
- Bilmes, J., & Zweig, G. (2002), "The graphical models toolkit: An open source software system for speech and time-series processing," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 3916–3919.
- Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, Springer.
- Blackwell, D., & MacQueen, J. B. (1973), "Ferguson distribution via Pólya urn schemes," *The Annals of Statistics* **1**, 353–355.
- Blei, D., Griffiths, T., & Jordan, M. (2010), "The nested Chinese restaurant process and Bayesian nonparametric inference of topic hierarchies," *Journal of the ACM* **57**(2), article 7.
- Blei, D., Griffiths, T., Jordan, M., & Tenenbaum, J. (2004), "Hierarchical topic models and the nested Chinese restaurant process," *Advances in Neural Information Processing Systems* **16**, 17–24.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003), "Latent Dirichlet allocation," *Journal of Machine Learning Research* **3**, 993–1022.
- Brants, T., Papat, A. C., Xu, P., Och, F. J., & Dean, J. (2007), "Large language models in machine translation," *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, Association for Computational Linguistics, pp. 858–867.
- Brill, E., & Moore, R. C. (2000), "An improved error model for noisy channel spelling correction," *Proceedings of the 38th Annual Meeting of Association for Computational Linguistics*, Association for Computational Linguistics, pp. 286–293.
- Brown, P., Desouza, P., Mercer, R., Pietra, V., & Lai, J. (1992), "Class-based *n*-gram models of natural language," *Computational Linguistics* **18**(4), 467–479.
- Brown, P. F., Cocke, J., Pietra, S. A. D., *et al.* (1990), "A statistical approach to machine translation," *Computational Linguistics* **16**(2), 79–85.
- Campbell, W. M., Sturim, D. E., & Reynolds, D. A. (2006), "Support vector machines using GMM supervectors for speaker verification," *Signal Processing Letters, IEEE* **13**(5), 308–311.
- Chen, K.-T., Liau, W.-W., Wang, H.-M., & Lee, L.-S. (2000), "Fast speaker adaptation using eigenspace-based maximum likelihood linear regression," *Proceedings of International Conference on Spoken Language Processing (ICSLP)*, pp. 742–745.

- Chen, S. F. (2009), "Shrinking exponential language models," in Proceedings of *Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, pp. 468–476.
- Chen, S. F., & Goodman, J. (1999), "An empirical study of smoothing techniques for language modeling," *Computer Speech & Language* **13**(4), 359–393.
- Chen, S., & Gopinath, R. (1999), "Model selection in acoustic modeling," Proceedings of European Conference on *Speech Communication and Technology (EUROSPEECH)*, pp. 1087–1090.
- Chesta, C., Siohan, O., & Lee, C.-H. (1999), "Maximum a posteriori linear regression for hidden Markov model adaptation," Proceedings of European Conference on *Speech Communication and Technology (EUROSPEECH)*, pp. 211–214.
- Chien, J.-T. (1999), "Online hierarchical transformation of hidden Markov models for speech recognition," *IEEE Transactions on Speech and Audio Processing* **7**(6), 656–667.
- Chien, J.-T. (2002), "Quasi-Bayes linear regression for sequential learning of hidden Markov models," *IEEE Transactions on Speech and Audio Processing* **10**(5), 268–278.
- Chien, J.-T. (2003), "Linear regression based Bayesian predictive classification for speech recognition," *IEEE Transactions on Speech and Audio Processing* **11**(1), 70–79.
- Chien, J.-T., & Chueh, C.-H. (2011), "Dirichlet class language models for speech recognition," *IEEE Transactions on Audio, Speech, and Language Processing* **19**(3), 482–495.
- Chien, J.-T., Huang, C.-H., Shinoda, K., & Furui, S. (2006), "Towards optimal Bayes decision for speech recognition," Proceedings of International Conference on *Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 45–48.
- Chien, J.-T., Lee, C.-H., & Wang, H.-C. (1997), "Improved Bayesian learning of hidden Markov models for speaker adaptation," Proceedings of International Conference on *Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 1027–1030.
- Chien, J. T., & Liao, G.-H. (2001), "Transformation-based Bayesian predictive classification using online prior evolution," *IEEE Transactions on Speech and Audio Processing* **9**(4), 399–410.
- Chien, J.-T., & Wu, M.-S. (2008), "Adaptive Bayesian latent semantic analysis," *IEEE Transactions on Audio, Speech, and Language Processing* **16**(1), 198–207.
- Chou, W., & Reichl, W. (1999), "Decision tree state tying based on penalized Bayesian information criterion," Proceedings of International Conference on *Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 345–348.
- Coccaro, N., & Jurafsky, D. (1998), "Towards better integration of semantic predictors in statistical language modeling," Proceedings of International Conference on *Spoken Language Processing (ICSLP)*, pp. 2403–2406.
- Cournapeau, D., Watanabe, S., Nakamura, A., & Kawahara, T. (2010), "Online unsupervised classification with model comparison in the variational Bayes framework for voice activity detection," *IEEE Journal of Selected Topics in Signal Processing* **4**(6), 1071–1083.
- Dahl, G. E., Yu, D., Deng, L., & Acero, A. (2012), "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," *IEEE Transactions on Audio, Speech and Language Processing* **20**(1), 30–42.
- Davis, S. B., & Mermelstein, P. (1980), "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing* **28**(4), 357–366.
- Dawid, A. P. (1981), "Some matrix-variate distribution theory: notational considerations and a Bayesian application," *Biometrika* **68**(1), 265–274.
- De Bruijn, N. G. (1970), *Asymptotic Methods in Analysis*, Dover Publications.

- Dehak, N., Kenny, P., Dehak, R., Dumouchel, P., & Ouellet, P. (2011), "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing* **19**(4), 788–798.
- Delcroix, M., Nakatani, T., & Watanabe, S. (2009), "Static and dynamic variance compensation for recognition of reverberant speech with dereverberation preprocessing," *IEEE Transactions on Audio, Speech, and Language Processing* **17**(2), 324–334.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1976), "Maximum likelihood from incomplete data via the EM algorithm," *Journal of Royal Statistical Society B* **39**, 1–38.
- Digalakis, V., & Neumeyer, L. (1996), "Speaker adaptation using combined transformation and Bayesian methods," *IEEE Transactions on Speech and Audio Processing* **4**, 294–300.
- Digalakis, V., Ritishev, D., & Neumeyer, L. (1995), "Speaker adaptation using constrained reestimation of Gaussian mixtures," *IEEE Transactions on Speech and Audio Processing* **3**, 357–366.
- Ding, N., & Ou, Z. (2010), "Variational nonparametric Bayesian hidden Markov model," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 2098–2101.
- Droppo, J., Acero, A., & Deng, L. (2002), "Uncertainty decoding with SPLICE for noise robust speech recognition," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 1, pp. I–57.
- Federico, M. (1996), "Bayesian estimation methods of n -gram language model adaptation," *Proceedings of International Conference on Spoken Language Processing (ICSLP)*, pp. 240–243.
- Ferguson, T. (1973), "A Bayesian analysis of some nonparametric problems," *The Annals of Statistics* **1**, 209–230.
- Fosler, E., & Morris, J. (2008), "Crandem systems: Conditional random field acoustic models for hidden Markov models," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 4049–4052.
- Fox, E. B., Sudderth, E. B., Jordan, M. I., & Willsky, A. S. (2008), "An HDP-HMM for systems with state persistence," *Proceedings of International Conference on Machine Learning (ICML)*, pp. 312–319.
- Fukunaga, K. (1990), *Introduction to Statistical Pattern Recognition*, Academic Press.
- Furui, S. (1981), "Cepstral analysis technique for automatic speaker verification," *IEEE Transactions on Acoustics, Speech and Signal Processing* **29**(2), 254–272.
- Furui, S. (1986), "Speaker independent isolated word recognition using dynamic features of speech spectrum," *IEEE Transactions on Acoustics, Speech and Signal Processing* **34**, 52–59.
- Furui, S. (2010), "History and development of speech recognition," in *Speech Technology*, F. Chen and K. Jokinen, eds., Springer, pp. 1–18.
- Furui, S., Maekawa, K., & H. Isahara, M. (2000), "A Japanese national project on spontaneous speech corpus and processing technology," *Proceedings of ASR'00*, pp. 244–248.
- Gales, M. (1998), "Maximum likelihood linear transformations for HMM-based speech recognition," *Computer Speech and Language* **12**, 75–98.
- Gales, M., Center, I., & Heights, Y. (2000), "Cluster adaptive training of hidden Markov models," *IEEE Transactions on Speech and Audio Processing* **8**(4), 417–428.
- Gales, M. J. F. (1999), "Semi-tied covariance matrices for hidden Markov models," *IEEE Transactions on Speech and Audio Processing* **7**(3), 272–281.
- Gales, M. J. F., & Woodland, P. C. (1996), Variance compensation within the MLLR framework, Technical Report 242, Cambridge University Engineering Department.

- Gales, M., Watanabe, S., & Fossler-Lussier, E. (2012), "Structured discriminative models for speech recognition," *IEEE Signal Processing Magazine* **29**(6), 70–81.
- Ganapathiraju, A., Hamaker, J., & Picone, J. (2004), "Applications of support vector machines to speech recognition," *IEEE Transactions on Signal Processing* **52**(8), 2348–2355.
- Gaussier, E., & Goutte, C. (2005), "Relation between PLSA and NMF and implications," Proceedings of the 28th Annual International ACM SIGIR Conference on *Research and Development in Information Retrieval*, ACM, pp. 601–602.
- Gauvain, J.-L., & Lee, C.-H. (1991), "Bayesian learning of Gaussian mixture densities for hidden Markov models," Proceedings of DARPA *Speech and Natural Language Workshop*, pp. 272–277.
- Gauvain, J.-L., & Lee, C.-H. (1994), "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains," *IEEE Transactions on Speech and Audio Processing* **2**, 291–298.
- Gelman, A., Carlin, J. B., Stern, H. S., *et al.* (2013), *Bayesian Data Analysis*, CRC Press.
- Geman, S., & Geman, D. (1984), "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**(1), 721–741.
- Genkin, A., Lewis, D. D., & Madigan, D. (2007), "Large-scale Bayesian logistic regression for text categorization," *Technometrics* **49**(3), 291–304.
- Ghahramani, Z. (1998), "Learning dynamic Bayesian networks," in *Adaptive Processing of Sequences and Data Structures*, Springer, pp. 168–197.
- Ghahramani, Z. (2004), "Unsupervised learning," *Advanced Lectures on Machine Learning*, pp. 72–112.
- Ghosh, J. K., Delampady, M., & Samanta, T. (2007), *An Introduction to Bayesian Analysis: Theory and Methods*, Springer.
- Gildea, D., & Hofmann, T. (1999), "Topic-based language models using EM," Proceedings of European Conference on *Speech Communication and Technology (EUROSPEECH)*, pp. 2167–2170.
- Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996), *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC Interdisciplinary Statistics.
- Gish, H., Siu, M.-h., Chan, A., & Belfield, W. (2009), "Unsupervised training of an HMM-based speech recognizer for topic classification," Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH), pp. 1935–1938.
- Glass, J. (2003), "A probabilistic framework for segment-based speech recognition," *Computer Speech & Language* **17**(2-3), 137–152.
- Goel, V., & Byrne, W. (2000), "Minimum Bayes-risk automatic speech recognition," *Computer Speech and Language* **14**, 115–135.
- Goldwater, S. (2007), Nonparametric Bayesian models of lexical acquisition, PhD thesis, Brown University.
- Goldwater, S., & Griffiths, T. (2007), "A fully Bayesian approach to unsupervised part-of-speech tagging," Proceedings of Annual Meeting of the Association of Computational Linguistics, pp. 744–751.
- Goldwater, S., Griffiths, T., & Johnson, M. (2009), "A Bayesian framework for word segmentation: Exploring the effects of context," *Cognition* **112**(1), 21–54.
- Goldwater, S., Griffiths, T. L., & Johnson, M. (2006), "Interpolating between types and tokens by estimating power-law generators," *Advances in Neural Information Processing Systems* **18**.

- Good, I. J. (1953), "The population frequencies of species and the estimation of populations," *Biometrika* **40**, 237–264.
- Grézl, F., Karafiát, M., Kontár, S., & Cernocký, J. (2007), "Probabilistic and bottle-neck features for LVCSR of meetings," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 757–760.
- Griffiths, T., & Ghahramani, Z. (2005), Infinite latent feature models and the Indian buffet process, Technical Report, Gatsby Unit.
- Griffiths, T., & Steyvers, M. (2004), "Finding scientific topics," in *Proceedings of the National Academy of Sciences*, **101** Suppl. 1, 5228–5235.
- Gunawardana, A., Mahajan, M., Acero, A., & Platt, J. C. (2005), "Hidden conditional random fields for phone classification," *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 1117–1120.
- Haeb-Umbach, R., & Ney, H. (1992), "Linear discriminant analysis for improved large vocabulary continuous speech recognition," *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 1, pp. 13–16.
- Hahm, S. J., Ogawa, A., Fujimoto, M., Hori, T., & Nakamura, A. (2012), "Speaker adaptation using variational Bayesian linear regression in normalized feature space," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 803–806.
- Hashimoto, K., Nankaku, Y., & Tokuda, K. (2009), "A Bayesian approach to hidden semi-Markov model based speech synthesis," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 1751–1754.
- Hashimoto, K., Zen, H., Nankaku, Y., Lee, A., & Tokuda, K. (2008), "Bayesian context clustering using cross valid prior distribution for HMM-based speech recognition," *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 936–939.
- Hashimoto, K., Zen, H., Nankaku, Y., Masuko, T., & Tokuda, K. (2009), "A Bayesian approach to HMM-based speech synthesis," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2009*, pp. 4029–4032.
- Hastings, W. K. (1970), "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika* **57**, 97–109.
- Heigold, G., Ney, H., Schluter, R., & Wiesler, S. (2012), "Discriminative training for automatic speech recognition: Modeling, criteria, optimization, implementation, and performance," *IEEE Signal Processing Magazine* **29**(6), 58–69.
- Hermansky, H. (1990), "Perceptual linear predictive (PLP) analysis of speech," *Journal of the Acoustic Society of America* **87**(4), 1738–1752.
- Hermansky, H., Ellis, D., & Sharma, S. (2000), "Tandem connectionist feature extraction for conventional HMM systems," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 1635–1638.
- Hinton, G., Deng, L., Yu, D., *et al.* (2012), "Deep neural networks for acoustic modeling in speech recognition," *IEEE Signal Processing Magazine* **29**(6), 82–97.
- Hinton, G., Osindero, S., & Teh, Y. (2006), "A fast learning algorithm for deep belief nets," *Neural Computation* **18**, 1527–1554.
- Hofmann, T. (1999a), "Probabilistic latent semantic analysis," *Proceedings of Uncertainty in Artificial Intelligence*, pp. 289–296.
- Hofmann, T. (1999b), "Probabilistic latent semantic indexing," *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 50–57.

- Hofmann, T. (2001), "Unsupervised learning by probabilistic latent semantic analysis," *Machine Learning* **42**(1-2), 177–196.
- Hori, T., & Nakamura, A. (2013), "Speech recognition algorithms using weighted finite-state transducers," *Synthesis Lectures on Speech and Audio Processing* **9**(1), 1–162.
- Hu, R., & Zhao, Y. (2007), "Knowledge-based adaptive decision tree state tying for conversational speech recognition," *IEEE Transactions on Audio, Speech, and Language Processing* **15**(7), 2160–2168.
- Huang, S., & Renals, S. (2008), "Unsupervised language model adaptation based on topic and role information in multiparty meeting," *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 833–836.
- Huang, X. D., Acero, A., & Hon, H. W. (2001), *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*, Prentice Hall.
- Huang, X. D., Ariki, Y., & Jack, M. A. (1990), *Hidden Markov Models for Speech Recognition*, Edinburgh University Press.
- Huo, Q., & Lee, C.-H. (1997), "On-line adaptive learning of the continuous density hidden Markov model based on approximate recursive Bayes estimate," *IEEE Transactions on Speech and Audio Processing* **5**(2), 161–172.
- Huo, Q., & Lee, C.-H. (2000), "A Bayesian predictive classification approach to robust speech recognition," *IEEE Transactions on Speech and Audio Processing* **8**, 200–204.
- Ishiguro, K., Yamada, T., Araki, S., Nakatani, T., & Sawada, H. (2012), "Probabilistic speaker diarization with bag-of-words representations of speaker angle information," *IEEE Transactions on Audio, Speech, and Language Processing* **20**(2), 447–460.
- Jansen, A., Dupoux, E., Goldwater, S., *et al.* (2013), "A summary of the 2012 JHU CLSP workshop on zero resource speech technologies and models of early language acquisition," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 8111–8115.
- Jelinek, F. (1976), "Continuous speech recognition by statistical methods," *Proceedings of the IEEE* **64**(4), 532–556.
- Jelinek, F. (1997), *Statistical Methods for Speech Recognition*, MIT Press.
- Jelinek, F., & Mercer, R. L. (1980), "Interpolated estimation of Markov source parameters from sparse data," *Proceedings of the Workshop on Pattern Recognition in Practice*, pp. 381–397.
- Ji, S., Xue, Y., & Carin, L. (2008), "Bayesian compressive sensing," *IEEE Transactions on Signal Processing* **56**(6), 2346–2356.
- Jiang, H., Hirose, K., & Huo, Q. (1999), "Robust speech recognition based on a Bayesian prediction approach," *IEEE Transactions on Speech and Audio Processing* **7**, 426–440.
- Jitsuhiro, T., & Nakamura, S. (2004), "Automatic generation of non-uniform HMM structures based on variational Bayesian approach," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 805–808.
- Joachims, T. (2002), "Learning to classify text using support vector machines: Methods, theory, and algorithms," *Computational Linguistics* **29**(4), 656–664.
- Jordan, M., Ghahramani, Z., Jaakkola, T., & Saul, L. (1999), "An introduction to variational methods for graphical models," *Machine Learning* **37**(2), 183–233.
- Juang, B.-H., & Rabiner, L. (1990), "The segmental K-means algorithm for estimating parameters of hidden Markov models," *IEEE Transactions on Acoustics, Speech and Signal Processing* **38**(9), 1639–1641.
- Juang, B., & Katagiri, S. (1992), "Discriminative learning for minimum error classification," *IEEE Transactions on Signal Processing* **40**(12), 3043–3054.

- Jurafsky, D. (2014), "From languages to information," <http://www.stanford.edu/class/cs124/lec/languagemodeling.pdf>.
- Jurafsky, D., & Martin, J. H. (2000), *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Prentice Hall.
- Kass, R. E., & Raftery, A. E. (1993), Bayes factors and model uncertainty, Technical Report 254, Department of Statistics, University of Washington.
- Kass, R. E., & Raftery, A. E. (1995), "Bayes factors," *Journal of the American Statistical Association* **90**(430), 773–795.
- Katz, S. (1987), "Estimation of probabilities from sparse data for the language model component of a speech recognizer," *IEEE Transactions on Acoustics, Speech, and Signal Processing* **35**(3), 400–401.
- Kawabata, T., & Tamoto, M. (1996), "Back-off method for n-gram smoothing based on binomial posteriori distribution," Proceedings of International Conference on *Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 1, pp. 192–195.
- Kenny, P. (2010), "Bayesian speaker verification with heavy tailed priors," Keynote Speech, Odyssey Speaker and Language Recognition Workshop.
- Kenny, P., Boulianne, G., Ouellet, P., & Dumouchel, P. (2007), "Joint factor analysis versus eigenchannels in speaker recognition," *IEEE Transactions on Audio, Speech, and Language Processing* **15**(4), 1435–1447.
- Kingsbury, B., Sainath, T. N., & Soltau, H. (2012), "Scalable minimum Bayes risk training of deep neural network acoustic models using distributed Hessian-free optimization," Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH), pp. 10–13.
- Kinnunen, T., & Li, H. (2010), "An overview of text-independent speaker recognition: from features to supervectors," *Speech Communication* **52**(1), 12–40.
- Kita, K. (1999), *Probabilistic Language Models*, University of Tokyo Press (in Japanese).
- Kneser, R., & Ney, H. (1995), "Improved backing-off for *m*-gram language modeling," Proceedings of International Conference on *Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 181–184.
- Kneser, R., Peters, J., & Klakow, D. (1997), "Language model adaptation using dynamic marginals," Proceedings of European Conference on Speech Communication and Technology (EUROSPEECH), pp. 1971–1974.
- Kolossa, D., & Haeb-Umbach, R. (2011), *Robust Speech Recognition of Uncertain or Missing Data: Theory and Applications*, Springer.
- Kubo, Y., Watanabe, S., Nakamura, A., & Kobayashi, T. (2010), "A regularized discriminative training method of acoustic models derived by minimum relative entropy discrimination," Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH), pp. 2954–2957.
- Kudo, T. (2005), "Mecab: Yet another part-of-speech and morphological analyzer," <http://mecab.sourceforge.net/>.
- Kuhn, R., & De Mori, R. (1990), "A cache-based natural language model for speech recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**(6), 570–583.
- Kuhn, R., Junqua, J., Ngyuen, P., & Niedzielski, N. (2000), "Rapid speaker adaptation in eigenvoice space," *IEEE Transactions on Speech and Audio Processing* **8**(6), 695–707.
- Kullback, S., & Leibler, R. A. (1951), "On information and sufficiency," *Annals of Mathematical Statistics* **22**(1), 79–86.

- Kwok, J. T.-Y. (2000), "The evidence framework applied to support vector machines," *IEEE Transactions on Neural Networks* **11**(5), 1162–1173.
- Kwon, O., Lee, T.-W., & Chan, K. (2002), "Application of variational Bayesian PCA for speech feature extraction," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 1, pp. 825–828.
- Lafferty, J., McCallum, A., & Pereira, F. (2001), "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," *Proceedings of International Conference on Machine Learning*, pp. 282–289.
- Lamel, L., Gauvain, J.-L., & Adda, G. (2002), "Lightly supervised and unsupervised acoustic model training," *Computer Speech & Language* **16**(1), 115–129.
- Lau, R., Rosenfeld, R., & Roukos, S. (1993), "Trigger-based language models: A maximum entropy approach," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 2, IEEE, pp. 45–48.
- Lee, C.-H., & Huo, Q. (2000), "On adaptive decision rules and decision parameter adaptation for automatic speech recognition," *Proceedings of the IEEE* **88**, 1241–1269.
- Lee, C.-H., Lin, C.-H., & Juang, B.-H. (1991), "A study on speaker adaptation of the parameters of continuous density hidden Markov models," *IEEE Transactions on Acoustics, Speech, and Signal Processing* **39**, 806–814.
- Lee, C.-Y. (2014), *Discovering linguistic structures in speech: models and applications*, PhD thesis, Massachusetts Institute of Technology.
- Lee, C.-Y., & Glass, J. (2012), "A nonparametric Bayesian approach to acoustic model discovery," *Proceedings of Annual Meeting of the Association for Computational Linguistics*, pp. 40–49.
- Lee, C.-Y., Zhang, Y., & Glass, J. (2013), "Joint learning of phonetic units and word pronunciations for ASR," *Proceedings of the 2013 Conference on Empirical Methods on Natural Language Processing (EMNLP)*, pp. 182–192.
- Lee, D. D., & Seung, H. S. (1999), "Learning the parts of objects by non-negative matrix factorization," *Nature* **401**(6755), 788–791.
- Leggetter, C. J., & Woodland, P. C. (1995), "Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models," *Computer Speech and Language* **9**, 171–185.
- Lewis, D. D. (1998), "Naive (Bayes) at forty: The independence assumption in information retrieval," *Proceedings of the 10th European Conference on Machine Learning*, Springer-Verlag, pp. 4–15.
- Liu, J. (1994), "The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem," *Journal of the American Statistical Association* **89**(427).
- Liu, J. S. (2008), *Monte Carlo Strategies in Scientific Computing*, Springer.
- Livescu, K., Glass, J. R., & Bilmes, J. (2003), "Hidden feature models for speech recognition using dynamic Bayesian networks," *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 2529–2532.
- MacKay, D. J. C. (1992a), "Bayesian interpolation," *Neural Computation* **4**(3), 415–447.
- MacKay, D. J. C. (1992b), "The evidence framework applied to classification networks," *Neural Computation* **4**(5), 720–736.
- MacKay, D. J. C. (1992c), "A practical Bayesian framework for back-propagation networks," *Neural Computation* **4**(3), 448–472.
- MacKay, D. J. C. (1995), "Probable networks and plausible predictions – a review of practical Bayesian methods for supervised neural networks," *Network: Computation in Neural Systems* **6**(3), 469–505.

- MacKay, D. J. C. (1997), Ensemble learning for hidden Markov models, Technical Report, Cavendish Laboratory, University of Cambridge.
- MacKay, D. J. C., & Peto, L. C. B. (1995), "A hierarchical Dirichlet language model," *Natural Language Engineering* **1**(3), 289–308.
- Maekawa, T., & Watanabe, S. (2011), "Unsupervised activity recognition with user's physical characteristics data," *Proceedings of International Symposium on Wearable Computers*, pp. 89–96.
- Mak, B., Kwok, J., & Ho, S. (2005), "Kernel eigenvoice speaker adaptation," *IEEE Transactions on Speech and Audio Processing* **13**(5), 984–992.
- Manning, C. D., & Schütze, H. (1999), *Foundations of Statistical Natural Language Processing*, MIT Press.
- Masataki, H., Sagisaka, Y., Hisaki, K., & Kawahara, T. (1997), "Task adaptation using MAP estimation in n -gram language modeling," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 783–786.
- Matsui, T., & Furui, S. (1994), "Comparison of text-independent speaker recognition methods using VQ-distortion and discrete/continuous HMMs," *IEEE Transactions on Speech and Audio Processing* **2**(3), 456–459.
- Matsumoto, Y., Kitauchi, A., Yamashita, T., *et al.* (1999), "Japanese morphological analysis system ChaSen version 2.0 manual," NAIST Technical Report.
- McCallum, A., & Nigam, K. (1998), "A comparison of event models for naive Bayes text classification," in *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI) Workshop on Learning for Text Categorization*, Vol. 752, pp. 41–48.
- McDermott, E., Hazen, T., Le Roux, J., Nakamura, A., & Katagiri, S. (2007), "Discriminative training for large-vocabulary speech recognition using minimum classification error," *IEEE Transactions on Audio, Speech, and Language Processing* **15**(1), 203–223.
- Meignier, S., Moraru, D., Fredouille, C., Bonastre, J.-F., & Besacier, L. (2006), "Step-by-step and integrated approaches in broadcast news speaker diarization," *Computer Speech & Language* **20**(2), 303–330.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953), "Equation of state calculations by fast computing machines," *Journal of Chemical Physics* **21**(6), 1087–1092.
- Minka, T. P. (2001), "Expectation propagation for approximate Bayesian inference," *Proceedings of Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 362–369.
- Mochihashi, D., Yamada, T., & Ueda, N. (2009), "Bayesian unsupervised word segmentation with nested Pitman–Yor language modeling," *Proceedings of Joint Conference of Annual Meeting of the ACL and International Joint Conference on Natural Language Processing of the AFNLP*, pp. 100–108.
- Mohri, M., Pereira, F., & Riley, M. (2002), "Weighted finite-state transducers in speech recognition," *Computer Speech and Language* **16**, 69–88.
- Moraru, D., Meignier, S., Besacier, L., Bonastre, J.-F., & Magrin-Chagnolleau, I. (2003), "The ELISA consortium approaches in speaker segmentation during the NIST 2002 speaker recognition evaluation," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 2, pp. 89–92.
- Mrva, D., & Woodland, P. C. (2004), "A PLSA-based language model for conversational telephone speech," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 2257–2260.

- Murphy, K. P. (2002), *Dynamic Bayesian networks: representation, inference and learning*, PhD thesis, University of California, Berkeley.
- Murphy, K. P., Weiss, Y., & Jordan, M. I. (1999), "Loopy belief propagation for approximate inference: An empirical study," *Proceedings of Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 467–475.
- Nadas, A. (1985), "Optimal solution of a training problem in speech recognition," *IEEE Transactions on Acoustics, Speech and Signal Processing* **33**(1), 326–329.
- Nakagawa, S. (1988), *Speech Recognition by Probabilistic Model*, Institute of Electronics, Information and Communication Engineers (IEICE) (in Japanese).
- Nakamura, A., McDermott, E., Watanabe, S., & Katagiri, S. (2009), "A unified view for discriminative objective functions based on negative exponential of difference measure between strings," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 1633–1636.
- Neal, R., & Hinton, G. (1998), "A view of the EM algorithm that justifies incremental, sparse, and other variants," *Learning in Graphical Models*, pp. 355–368.
- Neal, R. M. (1992), "Bayesian mixture modeling," *Proceedings of the Workshop on Maximum Entropy and Bayesian Methods of Statistical Analysis* **11**, 197–211.
- Neal, R. M. (1993), "Probabilistic inference using Markov chain Monte Carlo methods," Technical Report CRG-TR-93-1, Dept. of Computer Science, University of Toronto.
- Neal, R. M. (2000), "Markov chain sampling methods for Dirichlet process mixture models," *Journal of Computational and Graphical Statistics* **9**(2), 249–265.
- Neal, R. M. (2003), "Slice sampling," *Annals of Statistics* **31**, 705–767.
- Nefian, A. V., Liang, L., Pi, X., Liu, X., & Murphy, K. (2002), "Dynamic Bayesian networks for audio-visual speech recognition," *EURASIP Journal on Applied Signal Processing* **11**, 1274–1288.
- Neubig, G., Mimura, M., Mori, S., & Kawahara, T. (2010), "Learning a language model from continuous speech," *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 1053–1056.
- Ney, H., Essen, U., & Kneser, R. (1994), "On structuring probabilistic dependences in stochastic language modeling," *Computer Speech and Language* **8**, 1–38.
- Ney, H., Haeb-Umbach, R., Tran, B.-H., & Oerder, M. (1992), "Improvements in beam search for 10000-word continuous speech recognition," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 1, IEEE, pp. 9–12.
- Niesler, T., & Willett, D. (2002), "Unsupervised language model adaptation for lecture speech transcription," *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 1413–1416.
- Normandin, Y. (1992), "Hidden Markov models, maximum mutual information estimation, and the speech recognition problem," PhD thesis, McGill University, Montreal, Canada.
- Odell, J. J. (1995), *The use of context in large vocabulary speech recognition*, PhD thesis, Cambridge University.
- Ostendorf, M., & Singer, H. (1997), "HMM topology design using maximum likelihood successive state splitting," *Computer Speech and Language* **11**, 17–41.
- Paul, D. B., & Baker, J. M. (1992), "The design for the *Wall Street Journal*-based CSR corpus," *Proceedings of the Workshop on Speech and Natural Language*, Association for Computational Linguistics, pp. 357–362.
- Pettersen, S. (2008), *Robust speech recognition in the presence of additive noise*, PhD thesis, Norwegian University of Science and Technology.

- Pitman, J. (2002), "Poisson-Dirichlet and GEM invariant distributions for split-and-merge transformation of an interval partition," *Combinatorics, Probability and Computing* **11**, 501–514.
- Pitman, J. (2006), *Combinatorial Stochastic Processes*, Springer-Verlag.
- Pitman, J., & Yor, M. (1997), "The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator," *Annals of Probability* **25**(2), 855–900.
- Porteous, I., Newman, D., Ihler, A., *et al.* (2008), "Fast collapsed Gibbs sampling for latent Dirichlet allocation," Proceedings of the 14th ACM SIGKDD International Conference on *Knowledge Discovery and Data Mining*, pp. 569–577.
- Povey, D. (2003), Discriminative training for large vocabulary speech recognition, PhD thesis, Cambridge University.
- Povey, D., Burget, L., Agarwal, M., *et al.* (2010), "Subspace Gaussian mixture models for speech recognition," Proceedings of International Conference on *Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 4330–4333.
- Povey, D., Gales, M. J. F., Kim, D., & Woodland, P. C. (2003), "MMI-MAP and MPE-MAP for acoustic model adaptation," Proceedings of European Conference on *Speech Communication and Technology (EUROSPEECH)* **8**, 1981–1984.
- Povey, D., Ghoshal, A., Boulianne, G., *et al.* (2011), "The Kaldi speech recognition toolkit," Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU).
- Povey, D., Kanevsky, D., Kingsbury, B., *et al.* (2008), "Boosted MMI for model and feature-space discriminative training," Proceedings of International Conference on *Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 4057–4060.
- Povey, D., Kingsbury, B., Mangu, L., *et al.* (2005), "fMPE: Discriminatively trained features for speech recognition," Proceedings of International Conference on *Acoustics, Speech, and Signal Processing (ICASSP)* **1**, 961–964.
- Povey, D., & Woodland, P. C. (2002), "Minimum phone error and I-smoothing for improved discriminative training," Proceedings of International Conference on *Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 13–17.
- Povey, D., Woodland, P., & Gales, M. (2003), "Discriminative MAP for acoustic model adaptation," Proceedings of International Conference on *Acoustics, Speech, and Signal Processing (ICASSP)* **1**, 1–312.
- Price, P., Fisher, W., Bernstein, J., & Pallett, D. (1988), "The DARPA 1000-word resource management database for continuous speech recognition," Proceedings of International Conference on *Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 651–654.
- Rabiner, L. R., & Juang, B.-H. (1986), "An introduction to hidden Markov models," *IEEE ASSP Magazine* **3**(1), 4–16.
- Rabiner, L. R., & Juang, B.-H. (1993), *Fundamentals of Speech Recognition*, Vol. 14, PTR Prentice Hall.
- Rasmussen, C. E. (1999), "The infinite Gaussian mixture model," *Advances in Neural Information Processing Systems* **12**, 554–560.
- Rasmussen, C. E., & Williams, C. K. I. (2006), *Gaussian Processes for Machine Learning, Adaptive Computation and Machine Learning*, MIT Press.
- Reynolds, D., Quatieri, T., & Dunn, R. (2000), "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing* **10**(1-3), 19–41.
- Rissanen, J. (1984), "Universal coding, information, prediction and estimation," *IEEE Transactions on Information Theory* **30**, 629–636.
- Rodriguez, A., Dunson, D. B., & Gelfand, A. E. (2008), "The nested Dirichlet process," *Journal of the American Statistical Association* **103**(483), 1131–1154.

- Rosenfeld, R. (2000), "Two decades of statistical language modeling: Where do we go from here?," *Proceedings of the IEEE* **88**(8), 1270–1278.
- Sainath, T. N., Ramabhadran, B., Picheny, M., Nahamoo, D., & Kanevsky, D. (2011), "Exemplar-based sparse representation features: from TIMIT to LVCSR," *IEEE Transactions on Audio, Speech and Language Processing* **19**(8), 2598–2613.
- Saito, D., Watanabe, S., Nakamura, A., & Minematsu, N. (2012), "Statistical voice conversion based on noisy channel model," *IEEE Transactions on Audio, Speech, and Language Processing* **20**(6), 1784–1794.
- Salakhutdinov, R. (2009), Learning deep generative models, PhD thesis, University of Toronto.
- Salton, G., & Buckley, C. (1988), "Term-weighting approaches in automatic text retrieval," *Information Processing & Management* **24**(5), 513–523.
- Sanderson, C., Bengio, S., & Gao, Y. (2006), "On transforming statistical models for non-frontal face verification," *Pattern Recognition* **39**(2), 288–302.
- Sankar, A., & Lee, C.-H. (1996), "A maximum-likelihood approach to stochastic matching for robust speech recognition," *IEEE Transactions on Speech and Audio Processing* **4**(3), 190–202.
- Saon, G., & Chien, J.-T. (2011), "Some properties of Bayesian sensing hidden Markov models," *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 65–70.
- Saon, G., & Chien, J.-T. (2012a), "Bayesian sensing hidden Markov models," *IEEE Transactions on Audio, Speech, and Language Processing* **20**(1), 43–54.
- Saon, G., & Chien, J.-T. (2012b), "Large-vocabulary continuous speech recognition systems: A look at some recent advances," *IEEE Signal Processing Magazine* **29**(6), 18–33.
- Schalkwyk, J., Beeferman, D., Beaufays, F., *et al.* (2010), "'Your word is my command': Google search by voice: A case study," in *Advances in Speech Recognition*, Springer, pp. 61–90.
- Schlüter, R., Macherey, W., Müller, B., & Ney, H. (2001), "Comparison of discriminative training criteria and optimization methods for speech recognition," *Speech Communication* **34**(3), 287–310.
- Schultz, T., & Waibel, A. (2001), "Language-independent and language-adaptive acoustic modeling for speech recognition," *Speech Communication* **35**(1), 31–51.
- Schwarz, G. (1978), "Estimating the dimension of a model," *The Annals of Statistics* **6**(2), 461–464.
- Scott, S. (2002), "Bayesian methods for hidden Markov models," *Journal of the American Statistical Association* **97**(457), 337–351.
- Seide, F., Li, G., Chen, X., & Yu, D. (2011), "Conversational speech transcription using context-dependent deep neural networks," *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 437–440.
- Sethuraman, J. (1994), "A constructive definition of Dirichlet priors," *Statistica Sinica* **4**, 639–650.
- Shikano, K., Kawahara, T., Kobayashi, T., *et al.* (1999), *Japanese Dictation Toolkit – Free Software Repository for Automatic Speech Recognition*, <http://www.ar.media.kyoto-u.ac.jp/dictation/>.
- Shinoda, K. (2010), "Acoustic model adaptation for speech recognition," *IEICE Transactions on Information and Systems* **93**(9), 2348–2362.
- Shinoda, K., & Inoue, N. (2013), "Reusing speech techniques for video semantic indexing," *IEEE Signal Processing Magazine* **30**(2), 118–122.
- Shinoda, K., & Iso, K. (2001), "Efficient reduction of Gaussian components using MDL criterion for HMM-based speech recognition," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 869–872.

- Shinoda, K., & Lee, C.-H. (2001), "A structural Bayes approach to speaker adaptation," *IEEE Transactions on Speech and Audio Processing* **9**, 276–287.
- Shinoda, K., & Watanabe, T. (1996), "Speaker adaptation with autonomous model complexity control by MDL principle," *Proceedings of International Conference on Acoustic, Speech, and Signal Processing (ICASSP)*, pp. 717–720.
- Shinoda, K., & Watanabe, T. (1997), "Acoustic modeling based on the MDL criterion for speech recognition," *Proceedings of European Conference on Speech Communication and Technology (EUROSPEECH)*, Vol. 1, pp. 99–102.
- Shinoda, K., & Watanabe, T. (2000), "MDL-based context-dependent subword modeling for speech recognition," *Journal of the Acoustical Society of Japan (E)* **21**, 79–86.
- Shiota, S., Hashimoto, K., Nankaku, Y., & Tokuda, K. (2009), "Deterministic annealing based training algorithm for Bayesian speech recognition," *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 680–683.
- Siohan, O., Myrvoll, T. A., & Lee, C. H. (2002), "Structural maximum a posteriori linear regression for fast HMM adaptation," *Computer Speech and Language* **16**(1), 5–24.
- Siu, M.-h., Gish, H., Chan, A., Belfield, W., & Lowe, S. (2014), "Unsupervised training of an HMM-based self-organizing unit recognizer with applications to topic classification and keyword discovery," *Computer Speech & Language* **28**(1), 210–223.
- Somervuo, P. (2004), "Comparison of ML, MAP, and VB based acoustic models in large vocabulary speech recognition," *Proceedings of International Conference on Spoken Language Processing (ICSLP)*, pp. 830–833.
- Spiegelhalter, D. J., & Lauritzen, S. L. (1990), "Sequential updating of conditional probabilities on directed graphical structures," *Networks* **20**(5), 579–605.
- Sproat, R., Gale, W., Shih, C., & Chang, N. (1996), "A stochastic finite-state word-segmentation algorithm for Chinese," *Computational Linguistics* **22**(3), 377–404.
- Stenger, B., Ramesh, V., Paragios, N., Coetzee, F., & Buhmann, J. M. (2001), "Topology free hidden Markov models: Application to background modeling," *Proceedings of International Conference on Computer Vision (ICCV)*, Vol. 1, pp. 294–301.
- Stolcke, A., Ferrer, L., Kajarekar, S., Shriberg, E., & Venkataraman, A. (2005), "MLLR transforms as features in speaker recognition," *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 2425–2428.
- Stolcke, A., & Omohundro, S. (1993), "Hidden Markov model induction by Bayesian model merging," *Advances in Neural Information Processing Systems*, pp. 11–18, Morgan Kaufmann.
- Takahashi, J., & Sagayama, S. (1997), "Vector-field-smoothed Bayesian learning for fast and incremental speaker/telephone-channel adaptation," *Computer Speech and Language* **11**, 127–146.
- Takami, J., & Sagayama, S. (1992), "A successive state splitting algorithm for efficient allophone modeling," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 573–576.
- Tam, Y.-C., & Schultz, T. (2005), "Dynamic language model adaptation using variational Bayes inference," *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 5–8.
- Tam, Y.-C., & Schultz, T. (2006), "Unsupervised language model adaptation using latent semantic marginals," *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 2206–2209.
- Tamura, M., Masuko, T., Tokuda, K., & Kobayashi, T. (2001), "Adaptation of pitch and spectrum for HMM-based speech synthesis using MLLR," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 805–808.

- Tawara, N., Ogawa, T., Watanabe, S., & Kobayashi, T. (2012a), "Fully Bayesian inference of multi-mixture Gaussian model and its evaluation using speaker clustering," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 5253–5256.
- Tawara, N., Ogawa, T., Watanabe, S., Nakamura, A., & Kobayashi, T. (2012b), "Fully Bayesian speaker clustering based on hierarchically structured utterance-oriented Dirichlet process mixture model," *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 2166–2169.
- Teh, Y. W. (2006), "A hierarchical Bayesian language model based on Pitman–Yor processes," *Proceedings of International Conference on Computational Linguistics and Annual Meeting of the Association for Computational Linguistics*, pp. 985–992.
- Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2006), "Hierarchical Dirichlet processes," *Journal of the American Statistical Association* **101**(476), 1566–1581.
- Tipping, M. E. (2001), "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research* **1**, 211–244.
- Torbati, A. H. H. N., Picone, J., & Sobel, M. (2013), "Speech acoustic unit segmentation using hierarchical Dirichlet processes," *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 637–641.
- Ueda, N., & Ghahramani, Z. (2002), "Bayesian model search for mixture models based on optimizing variational bounds," *Neural Networks* **15**, 1223–1241.
- Valente, F. (2006), "Infinite models for speaker clustering," *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 1329–1332.
- Valente, F., Motlicek, P., & Vijayasenan, D. (2010), "Variational Bayesian speaker diarization of meeting recordings," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 4954–4957.
- Valente, F., & Wellekens, C. (2003), "Variational Bayesian GMM for speech recognition," *Proceedings of European Conference on Speech Communication and Technology (EUROSPEECH)*, pp. 441–444.
- Valente, F., & Wellekens, C. (2004a), "Variational Bayesian feature selection for Gaussian mixture models," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 1, pp. 513–516.
- Valente, F., & Wellekens, C. (2004b), "Variational Bayesian speaker clustering," *Proceedings of ODYSSEY The Speaker and Language Recognition Workshop*, pp. 207–214.
- Vapnik, V. (1995), *The Nature of Statistical Learning Theory*, Springer-Verlag.
- Veselý, K., Ghoshal, A., Burget, L., & Povey, D. (2013), "Sequence-discriminative training of deep neural networks," *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 2345–2349.
- Villalba, J., & Brümmer, N. (2011), "Towards fully Bayesian speaker recognition: Integrating out the between-speaker covariance," *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 505–508.
- Vincent, E., Barker, J., Watanabe, S., *et al.* (2013), "The second 'CHiME' speech separation and recognition challenge: An overview of challenge systems and outcomes," *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 162–167.
- Viterbi, A. J. (1967), "Error bounds for convolutional codes and an asymptotically optimal decoding algorithm," *IEEE Transactions on Information Theory* **IT-13**, 260–269.
- Wallach, H. M. (2006), "Topic modeling: beyond bag-of-words," *Proceedings of International Conference on Machine Learning*, pp. 977–984.

- Watanabe, S., & Chien, J. T. (2012), "Tutorial: Bayesian learning for speech and language processing," *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.
- Watanabe, S., Minami, Y., Nakamura, A., & Ueda, N. (2002), "Application of variational Bayesian approach to speech recognition," *Advances in Neural Information Processing Systems*.
- Watanabe, S., Minami, Y., Nakamura, A., & Ueda, N. (2004), "Variational Bayesian estimation and clustering for speech recognition," *IEEE Transactions on Speech and Audio Processing* **12**, 365–381.
- Watanabe, S., & Nakamura, A. (2004), "Acoustic model adaptation based on coarse-fine training of transfer vectors and its application to a speaker adaptation task," *Proceedings of International Conference on Spoken Language Processing (ICSLP)*, pp. 2933–2936.
- Watanabe, S., & Nakamura, A. (2006), "Speech recognition based on Student's t-distribution derived from total Bayesian framework," *IEICE Transactions on Information and Systems* **E89-D**, 970–980.
- Watanabe, S., & Nakamura, A. (2009), "On-line adaptation and Bayesian detection of environmental changes based on a macroscopic time evolution system," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 4373–4376.
- Watanabe, S., Nakamura, A., & Juang, B. (2011), "Bayesian linear regression for hidden Markov model based on optimizing variational bounds," *Proceedings of IEEE Workshop on Machine Learning for Signal Processing*, pp. 1–6.
- Watanabe, S., Nakamura, A., & Juang, B.-H. (2013), "Structural Bayesian linear regression for hidden Markov models," *Journal of Signal Processing Systems*, 1–18.
- Wegmann, S., McAllaster, D., Orloff, J., & Peskin, B. (1996), "Speaker normalization on conversational telephone speech," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 339–341.
- Winn, J., & Bishop, C. (2006), "Variational message passing," *Journal of Machine Learning Research* **6**(1), 661.
- Witten, I. H., & Bell, T. C. (1991), "The zero-frequency problem: estimating the probabilities of novel events in adaptive text compression," *IEEE Transactions on Information Theory* **37**, 1085–1094.
- Wooters, C., Fung, J., Peskin, B., & Anguera, X. (2004), "Towards robust speaker segmentation: The ICSI-SRI fall 2004 diarization system," in *RT-04F Workshop*, Vol. 23.
- Wooters, C., & Huijbregts, M. (2008), "The ICSI RT07s speaker diarization system," in *Multimodal Technologies for Perception of Humans*, Springer, pp. 509–519.
- Yamagishi, J., Kobayashi, T., Nakano, Y., Ogata, K., & Isogai, J. (2009), "Analysis of speaker adaptation algorithms for HMM-based speech synthesis and a constrained SMAPLR adaptation algorithm," *IEEE Transactions on Audio, Speech, and Language Processing* **17**(1), 66–83.
- Yaman, S., Chien, J.-T., & Lee, C.-H. (2007), "Structural Bayesian language modeling and adaptation," *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 2365–2368.
- Yedidia, J. S., Freeman, W. T., & Weiss, Y. (2003), "Understanding belief propagation and its generalizations," *Exploring Artificial Intelligence in the New Millennium* **8**, 236–239.
- Young, S., Evermann, G., Gales, M., *et al.* (2006), "The HTK book (for HTK version 3.4)," Cambridge University Engineering Department.
- Young, S. J., Odell, J. J., & Woodland, P. C. (1994), "Tree-based state tying for high accuracy acoustic modelling," *Proceedings of the Workshop on Human Language Technology*, pp. 307–312.

- Yu, K., & Gales, M. J. F. (2006), "Incremental adaptation using Bayesian inference," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 217–220.
- Zhang, Y., & Glass, J. R. (2009), "Unsupervised spoken keyword spotting via segmental DTW on Gaussian posteriorgrams," *Proceedings of IEEE Automatic Speech Recognition & Understanding Workshop (ASRU)*, pp. 398–403.
- Zhang, Y., Liu, P., Chien, J.-T., & Soong, F. (2009), "An evidence framework for Bayesian learning of continuous-density hidden Markov models," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 3857–3860.
- Zhao, X., Dong, Y., Zhao, J., *et al.* (2009), "Variational Bayesian joint factor analysis for speaker verification," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 4049–4052.
- Zhou, B., & Hansen, J. H. (2000), "Unsupervised audio stream segmentation and clustering via the Bayesian information criterion," *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, pp. 714–717.
- Zweig, G., & Nguyen, P. (2009), "A segmental CRF approach to large vocabulary continuous speech recognition," *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 152–157.
- Zweig, G., & Russell, S. (1998), "Speech recognition with dynamic Bayesian networks," *Proceedings of the National Conference on Artificial Intelligence*, pp. 173–180.