"Discrimination of speech from nonspeech based on multiscale spectro-temporal modulations," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 3, pp. 920–930, 2006.

[6] T. Ng, B. Zhang, L. Nguyen, S. Matsoukas, X. Zhou, N. Mesgarani, K. Vesel ` y, and P. Matejka, "Developing a speech activity detection system for the DARPA RATS program." in *Proc. Interspeech*, 2012.

[7] N. Ryant, M. Liberman, and J. Yuan, "Speech activity detection on YouTube using deep neural networks," in *Proc. Interspeech*, 2013.

[8] D. M. Green and J. A. Swets, *Signal detection theory and psychophysics*. Wiley New York, 1966, vol. 1974.

[9] D. Freeman, G. Cosier, C. Southcott, and I. Boyd, "The voice activity detector for the Pan-European digital cellular mobile telephone service," in *Proc. ICASSP*, 1989.

[10] N. Kitaoka, K. Yamamoto, T. Kusamizu, S. Nakagawa, T. Yamada, S. Tsuge, C. Miyajima, T. Nishiura, M. Nakayama, and Y. e. a. Denda, "Development of VAD evaluation framework CENSREC-1-C and investigation of relationship between VAD and speech recognition performance," in *Proc. ASRU*, 2007.

[11] F. Beritelli, S. Casale, and G. Ruggeri, "A psychoacoustic auditory model to evaluate the performance of a voice activity detector," *Signal processing*, vol. 80, no. 7, pp. 1393–1397, 2000.

[12] C. Breslin, M. Gasic, M. Henderson, D. Kim, M. Szummer, B. Thomson, P. Tsiakoulis, and S. Young, "Continuous ASR for flexible incremental dialogue," in *Proc. ICASSP*, 2013.

[13] T. V. Pham, C. T. Tang, and M. Stadtschnitzer, "Using artificial neural network for robust voice activity detection under adverse conditions," in *International Conference on Computing and Communication Technologies RIVF'09*, 2009.

[14] J. P. Egan, *Signal detection theory and ROC-analysis*. Academic Press, 1975.

[15] J. A. Swets, "The relative operating characteristic in psychology," *Science*, vol. 182, no. 4116, pp. 990–1000, 1973.

[16] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET curve in assessment of detection task performance," in *Proc.EuroSpeech*, 1997.

[17] D. Reich, F. Putze, D. Heger, J. Ijsselmuiden, R. Stiefelhagen, and T. Schultz, "A real-time speech command detector for a smart control room," in *Proc. Interspeech*, 2011.

[18] Mehta, C. K. Pham, and C. E. Siong, "Linear dynamic models for voice activity detection." in *Proc. Interspeech*, 2011.

[19] A. V. Ivanov and G. Riccardi, "Automatic turn segmentation in spoken conversations." in *Proc. Interspeech*, 2010.

[20] D. B. Dean, S. Sridharan, R. J. Vogt, and M. W. Mason, "The QUT-NOISE-TIMIT corpus for the evaluation of voice activity detection algorithms," in *Proc. Interspeech*, 2010.

[21] H. Ghaemmaghami, B. J. Baker, R. J. Vogt, and S. Sridharan, "Noise robust voice activity detection using features extracted from the time-domain autocorrelation function," in *Proc. Interspeech*, 2010.

[22] Y. Mamiya, J. Yamagishi, O. Watts, R. A. Clark, S. King, and A. Stan, "Lightly supervised GMM VAD to use audiobook for speech synthesiser," in *Proc. ICASSP*, 2013.

[23] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.

[24] T. X. He, Y. C. Fan, Y. M. Qian, T. Tan, and K. Yu, "Reshaping deep neural network for fast decoding by node-pruning," in *Proc. ICASSP*, 2014.