

-
- ITU-T. Recommendation G. 711. *Pulse Code Modulation (PCM) of voice frequencies*, 1988.
- Józefowicz, Rafal, Vinyals, Oriol, Schuster, Mike, Shazeer, Noam, and Wu, Yonghui. Exploring the limits of language modeling. *CoRR*, abs/1602.02410, 2016. URL <http://arxiv.org/abs/1602.02410>.
- Juang, Biing-Hwang and Rabiner, Lawrence. Mixture autoregressive hidden Markov models for speech signals. *IEEE Trans. Acoust. Speech Signal Process.*, pp. 1404–1413, 1985.
- Kameoka, Hirokazu, Ohishi, Yasunori, Mochihashi, Daichi, and Le Roux, Jonathan. Speech analysis with multi-kernel linear prediction. In *Spring Conference of ASJ*, pp. 499–502, 2010. (in Japanese).
- Karaali, Orhan, Corrigan, Gerald, Gerson, Ira, and Massey, Noel. Text-to-speech conversion with neural networks: A recurrent TDNN approach. In *Eurospeech*, pp. 561–564, 1997.
- Kawahara, Hideki, Masuda-Katsuse, Ikuyo, and de Cheveigné, Alain. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based f_0 extraction: possible role of a repetitive structure in sounds. *Speech Commn.*, 27:187–207, 1999.
- Kawahara, Hideki, Estill, Jo, and Fujimura, Osamu. Aperiodicity extraction and control using mixed mode excitation and group delay manipulation for a high quality speech analysis, modification and synthesis system STRAIGHT. In *MAVEBA*, pp. 13–15, 2001.
- Law, Edith and Von Ahn, Luis. Input-agreement: a new mechanism for collecting data using human computation games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1197–1206. ACM, 2009.
- Maia, Ranniery, Zen, Heiga, and Gales, Mark J. F. Statistical parametric speech synthesis with joint estimation of acoustic and excitation model parameters. In *ISCA SSW7*, pp. 88–93, 2010.
- Morise, Masanori, Yokomori, Fumiya, and Ozawa, Kenji. WORLD: A vocoder-based high-quality speech synthesis system for real-time applications. *IEICE Trans. Inf. Syst.*, E99-D(7):1877–1884, 2016.
- Moulines, Eric and Charpentier, Francis. Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Commn.*, 9:453–467, 1990.
- Muthukumar, P. and Black, Alan W. A deep learning approach to data-driven parameterizations for statistical parametric speech synthesis. *arXiv:1409.8558*, 2014.
- Nair, Vinod and Hinton, Geoffrey E. Rectified linear units improve restricted Boltzmann machines. In *ICML*, pp. 807–814, 2010.
- Nakamura, Kazuhiro, Hashimoto, Kei, Nankaku, Yoshihiko, and Tokuda, Keiichi. Integration of spectral feature extraction and modeling for HMM-based speech synthesis. *IEICE Trans. Inf. Syst.*, E97-D(6):1438–1448, 2014.
- Palaz, Dimitri, Collobert, Ronan, and Magimai-Doss, Mathew. Estimating phoneme class conditional probabilities from raw speech signal using convolutional neural networks. In *Interspeech*, pp. 1766–1770, 2013.
- Peltonen, Sari, Gabbouj, Moncef, and Astola, Jaakko. Nonlinear filter design: methodologies and challenges. In *IEEE ISPA*, pp. 102–107, 2001.
- Poritz, Alan B. Linear predictive hidden Markov models and the speech signal. In *ICASSP*, pp. 1291–1294, 1982.
- Rabiner, Lawrence and Juang, Biing-Hwang. *Fundamentals of Speech Recognition*. PrenticeHall, 1993.
- Sagisaka, Yoshinori, Kaiki, Nobuyoshi, Iwahashi, Naoto, and Mimura, Katsuhiko. ATR ν -talk speech synthesis system. In *ICSLP*, pp. 483–486, 1992.