
Gaffney and Steve Crossan for helping to manage the project, Faith Mackinder for help with preparing the blogpost, James Besley for legal support and Demis Hassabis for managing the project and his inputs.

REFERENCES

- Agiomyrgiannakis, Yannis. Vocode the vocoder and applications in speech synthesis. In *ICASSP*, pp. 4230–4234, 2015.
- Bishop, Christopher M. Mixture density networks. Technical Report NCRG/94/004, Neural Computing Research Group, Aston University, 1994.
- Chen, Liang-Chieh, Papandreou, George, Kokkinos, Iasonas, Murphy, Kevin, and Yuille, Alan L. Semantic image segmentation with deep convolutional nets and fully connected CRFs. In *ICLR*, 2015. URL <http://arxiv.org/abs/1412.7062>.
- Chiba, Tsutomu and Kajiyama, Masato. *The Vowel: Its Nature and Structure*. Tokyo-Kaiseikan, 1942.
- Dudley, Homer. Remaking speech. *The Journal of the Acoustical Society of America*, 11(2):169–177, 1939.
- Dutilleul, Pierre. An implementation of the “algorithme à trous” to compute the wavelet transform. In Combes, Jean-Michel, Grossmann, Alexander, and Tchamitchian, Philippe (eds.), *Wavelets: Time-Frequency Methods and Phase Space*, pp. 298–304. Springer Berlin Heidelberg, 1989.
- Fan, Yuchen, Qian, Yao, and Xie, Feng-Long, Soong Frank K. TTS synthesis with bidirectional LSTM based recurrent neural networks. In *Interspeech*, pp. 1964–1968, 2014.
- Fant, Gunnar. *Acoustic Theory of Speech Production*. Mouton De Gruyter, 1970.
- Garofolo, John S., Lamel, Lori F., Fisher, William M., Fiscus, Jonathon G., and Pallett, David S. DARPA TIMIT acoustic-phonetic continuous speech corpus CD-ROM. NIST speech disc 1-1.1. *NASA STI/Recon technical report*, 93, 1993.
- Gonzalvo, Xavi, Tazari, Siamak, Chan, Chun-an, Becker, Markus, Gutkin, Alexander, and Silen, Hanna. Recent advances in Google real-time HMM-driven unit selection synthesizer. In *Interspeech*, 2016. URL <http://research.google.com/pubs/pub45564.html>.
- He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, and Sun, Jian. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.
- Hochreiter, S. and Schmidhuber, J. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, 1997.
- Holschneider, Matthias, Kronland-Martinet, Richard, Morlet, Jean, and Tchamitchian, Philippe. A real-time algorithm for signal analysis with the help of the wavelet transform. In Combes, Jean-Michel, Grossmann, Alexander, and Tchamitchian, Philippe (eds.), *Wavelets: Time-Frequency Methods and Phase Space*, pp. 286–297. Springer Berlin Heidelberg, 1989.
- Hoshen, Yedid, Weiss, Ron J., and Wilson, Kevin W. Speech acoustic modeling from raw multi-channel waveforms. In *ICASSP*, pp. 4624–4628. IEEE, 2015.
- Hunt, Andrew J. and Black, Alan W. Unit selection in a concatenative speech synthesis system using a large speech database. In *ICASSP*, pp. 373–376, 1996.
- Imai, Satoshi and Furuichi, Chieko. Unbiased estimation of log spectrum. In *EURASIP*, pp. 203–206, 1988.
- Itakura, Fumitada. Line spectrum representation of linear predictor coefficients of speech signals. *The Journal of the Acoust. Society of America*, 57(S1):S35–S35, 1975.
- Itakura, Fumitada and Saito, Shuzo. A statistical method for estimation of speech spectral density and formant frequencies. *Trans. IEICE*, J53A:35–42, 1970.