1. Hello Edge: KWS on Microcontrollers

Good overview of the KWS problem. More microcontroller focused (research on ARM chips, for example). Overview of different approaches (DNN, CNN, RNN, CRNN, DS-CNN). Includes the “deployment on micro-controller” example.

1. Robust keyword spotting in embedded systems, Ekaterina Chuikova (Lit Review)

Helpful background and overview of the keyword spotting (KWS) problem, and the key considerations for a successful KWS system. (Low latency, space constraints, accuracy.)

1. Compressing Deep Neural Networks using a Rank-Constrained Topology, Nakkiran et al.

This is the low-latency CNN that is implemented in our repository. This was implemented by Tensorflow. Worth reading to understand different ways of reducing the amount of FLOPs taken.

1. End To End Architecture for Keyword Spotting and Voice Activity Detection, Lengerich and Awni

Recurrent Network using [CTC](https://towardsdatascience.com/intuitively-understanding-connectionist-temporal-classification-3797e43a86c). They use a CNN into deep RNN. Main advantage: does not require frame-aligned labels. CTC computes score over all possible alignments. Speaks to improvement of the robustness if you give noisy data.

1. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling, Bai. et al

Questions the common association of sequence modeling with RNNs. Establishes foundation for Temporal Convolutional Networks (TCNs), which use CNN architecture on Sequence Models, achieving state of the art results. Linked to next paper: Bhattacharajee (sp.?)

Good background for causal convolutions, dilated convolutions and otherfoundations of TCNs.

1. Small-footprint Keyword Spotting using Deep Neural Networks, Chen, et al

Earlier paper for deep neural network keyword spotting, focuses on comparison with Hidden Markov Models. Helpful for posterior smoothing algorithm (i.e, how do you smooth the noisy outputs from the neural network?)

1. End to End Models with Auditory Attention in multi-channel keyword spotting

How attention (encoder-decoder) is used in keyword-spotting. Split into (1) Attention section, (2) Seq-to-seq training, (3) decoding smoothing.

1. Deep Residual Learning for Image Recognition

Not directly related to speech, but computer vision. Background around ResNet, one of the leading architectures for Convolutional Neural Networks. Explaining how residual networks improve CNN performance across the board. Deep representations 🡪 improvement on various tasks such as object detection.

1. Temporal Convolution for Real-Time Keyword Spotting on Mobile Devices

How TCN is implemented for keyword spotting. Insight; we can see TCN as 1D-Convolutions along the time-axis. They adopt Res-Net as their base CNN. They have comparisons between architectures in the paper. They also have an implementation. (<https://github.com/hyperconnect/TC-ResNet)>