P.p1 {margin: 0.0px 0.0px 0.0px 0.0px; font:

Design, Architecture

p. p1 {margin: 0. 0px 0. 0px 0.

0px 0. 0px; font: 10. 0px Helvetica}p. p2 {margin: 0. 0px 0.

0px 0. 0px 0. 0px; font: 12. 0px Helvetica}span. s1 {color: #002486} Speech recognition is the translation, throughsome methodologies, of human speech into textby computers. In this research review we examinethree di erent methods that are used inspeech recognition field and we investigate theaccuracy they succeed in di erent data sets. Weanalyze the state-of-art deep neural networks, thathave evolved into complex architectures and theyachieve significant results in many cases. Afterward, we explain convolutional neural networksand we explore their dynamic in this field.

Finally, we present the recent research in highwaydeep neural networks that seem to be more flexiblefor resource constrained platforms. Overall, we critically try to compare these methods andshow their strengths and limitations. We concludethat each method has its advantages butalso has its weaknesses and we use them for differentpurposes. I. Introduction Machine Learning (ML) is a field of computer science thatgives the computers the ability to learn through di erentalgorithms and techniques without being programmed. Automaticspeech recognition (ASR) is closely related withML because it uses methodologies and procedures of ML1 , 2 , 3 .

ASR has been around for decades but it was notuntil recently that there was a tremendous developmentbecause of the advances in both machine learning methodsand computer hardware. New ML techniques madespeech recognition accurate enough to be useful outsideof carefully controlled

environments and so it could easilybe deployed in many electronic devices nowadays (i. e. computers, smart-phones).

Speech is the most important mode of communicationbetween human beings and so from the early part of theprevious century, e orts have been made in order to makecomputers do what only humans could perceive. Researchhas been conducted through the past five decades and themain reason was the desire of making tasks automated usingmachines 2 . Many motivations from the field of machinelearning and the perspective of probabilistic modeling andreasoning to the neural a ected the researchers and helpedto advance ASR. The first single advance in the history of ASR occurredat the early of 50’s with the introduction of the expectationmaximization(EM) algorithm for training Hidden MarkovModels (HMMs).

The EM technique gave the possibility todevelop the first speech recognition systems using GaussianMixture Models (GMMs). Despite all the advantages of theGMMs, they are statistically ine cient for modeling datathat lie on or near a nonlinear manifold in the data space. This problem could be solved by artificial neural networks. Most speech recognition systems use neural network andhidden Markov model (NN/ HMM) hybrid architecture, firstinvestigated in the early 1990s 4 . However computer hardwaredid not allow us to train our data with more complexnetworks such as deep neural networks (DNNs) until theearly of 2000s. Over the last years the improvement of computerhardware and the invention of new machine learningalgorithms made possible the training for DNNs. DNNswith many

hidden layers have been showed to outperformGMMs on a variety of speech recognition benchmarks 5 . Other more complex neural architectures such as recurrentneural networks with long short-term memory units(LSTM-RNNs) and convolutional neural networks (CNNs)seem to have their benefits and applications.