ASA Spring Meeting , Invited Talk, 1988.5.16~20, Seattle, USA,

Speech Recognition Using Time-Delay Neural Networks

A. Waibel *
ATR Interpreting Telephony Research Laboratories

January 19, 1988

Abstract

We present a Time Delay Neural Network (TDNN) approach to speech recognition which is characterized by two important properties: 1.) Using multilayer arrangements of simple computing units, a TDNN can represent arbitrary nonlinear classification decision surfaces that are learned automatically using error back-propagation. 2.) The time-delay arrangement enables the network to discover acoustic-phonetic features and the temporal relationships between them independent of position in time and hence not blurred by temporal shifts in the input. We compare TDNNs with the currently most popular technique in speech recognition, Hidden Markov Models (HMM). Extensive performance evaluation shows that the TDNN recognizes voiced stops extracted from varying phonetic contexts at an error rate four times lower (1.5 % vs. 6.3 %) than the best of our HMMs. To perform this task, the TDNN "invented" well-known acoustic phonetic features (e.g., F2-rise, F2-fall, vowel-onset) as useful abstractions. It also developed alternate internal representations to link different acoustic realizations to the same concept. TDNNs trained for other phonetic classes achieve similar high levels of performance. We discuss the integration of such smaller networks into large phonetic nets and propose strategies for the design of neural network based large vocabulary speech recognition systems.

*This work was performed in collaboration with T. Hanazawa and K. Shikano of ATR, G. Hinton of the University of Toronto and Canadian Institute for Advanced Research and K. Lang of Carnegie-Mellon University

USA-Japan Joint Acoustical Society Meeting, 1988.11.14, Hawaii, USA.

Incremental Learning of Large Phonetic Neural Networks from Smaller Subnets. A. Waibel (ATR Interpreting Telephony Research Laboratories, Twin 21 MID Tower, 2-1-61 Shiromi, Osaka, 540, Japan)

Time Delay Neural Networks (TDNNs) [Waibel et al., ATR-TR-0006 (1987)] have recently been shown to achieve excellent recognition performance for difficult phonetic discrimination tasks (e.g., voiced stops). This was achieved in part by the TDNNs' ability to not only activate the correct output category, but also to inhibit all incorrect outputs. The disadvantage of this property is that training larger networks with many more output categories (e.g., all consonants) becomes non-trivial as all categories have to be incorporated in the learning process requiring excessive amounts of training. Several techniques are presented that overcome this problem by exploiting the hidden structure of previously learned smaller neural nets to train larger nets incrementally in comparatively short training runs. Experimental results show that the resulting larger networks aimed at stop consonants, at voiced stops and nasals and at all consonants achieve recognition scores (95% to 99%) as high as the smaller subnetworks from which they were constructed.

Suggested for Special Session on Speech Communication, Human-Machine Interaction

Technical Committee: Speech Communication

Method of Presentation: Prefer Lecture but Willing to Give as Poster

(PACS) Subject Classification Number: 43.72.Ne

Telephone Number: 412-268-7674 (Secretary, Karen Olack)

Send Acceptance or Rejection Notice to:

A. Waibel, Computer Science Department, Carnegie-Mellon University, Pittsburgh, PA 15213