Snowbird Conference, 1988. 4. 6 ~ 9.

Speech Recognition Using Time-Delay Neural Networks
Invited Talk
Snowbird, 1988

A. Waibel *
ATR Interpreting Telephony Research Laboratories

December 3, 1987

Abstract

In this talk we present a Time Delay Neural Network (TDNN) approach to phoneme recognition which is characterized by two important properties: 1.) Using a 3 layer arrangement of simple computing units, it can represent arbitrary nonlinear decision surfaces. The TDNN learns these decision surfaces 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 our TDNNs with the currently most popular technique in speech recognition: Hidden Markov Models (HMM). Several HMMs were trained to perform the same task as our TDNN, i.e., the speaker-dependent recognition of the phonemes "B", "D", and "G" extracted from varying phonetic contexts. Extensive performance evaluation shows that the TDNN achieved a recognition rate of 98.5 % correct, compared to only 93.7 % for the best of our HMMs. We show that 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. We will conclude by discussing the implications of this result for the design of larger scale 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