

SEE ME, HEAR ME: INTEGRATING AUTOMATIC SPEECH RECOGNITION AND LIP-READING

*Paul Duchnowski*¹

*Uwe Meier*¹

Alex Waibel^{1,2}

¹University of Karlsruhe, Karlsruhe, Germany

²Carnegie Mellon University, Pittsburgh PA, USA

ABSTRACT

Recent work on integration of visual informa-
tion with acoustic speech for lip-

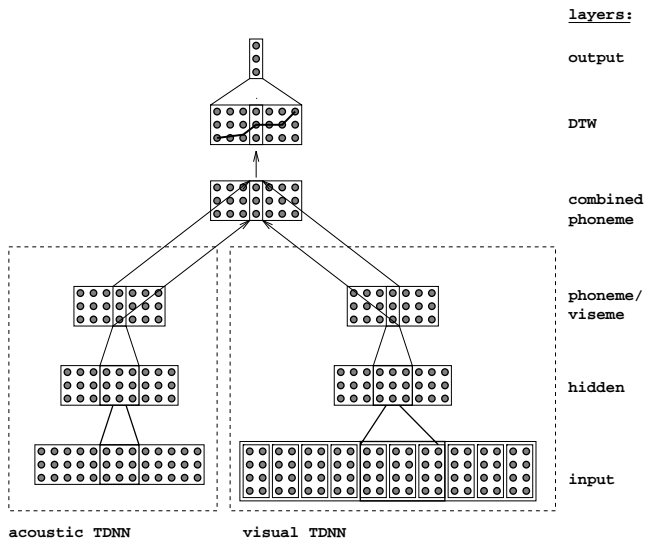


Figure 1. Original recognition network architecture (Net-P).

data, respectively. Weighted sum of the phoneme and combined
 in activations are entered in the combined
 path

Input	Visual Count	Parameter data set	Val. Accuracy (%)
84	55	ml-2	
53	45	ml-10	
38	44		

a repre-

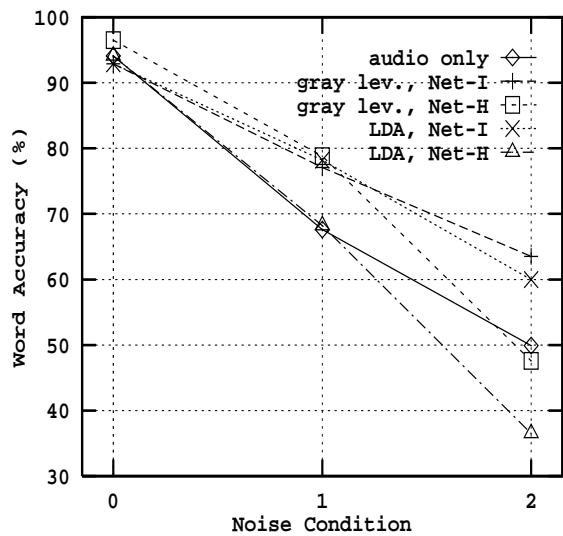


Figure 4. Combination results for Net-I and Net-H.

Comparison of different net structures yields more equiv-
 three are clearly capable of improving
 on how